

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :  
C12N 15/31, C07K 14/315, 16/12, C12Q  
1/68

A2

(11) International Publication Number: WO 98/50555

(43) International Publication Date: 12 November 1998 (12.11.98)

(21) International Application Number: PCT/US98/08985

(22) International Filing Date: 4 May 1998 (04.05.98)

(30) Priority Data:

60/044,031	6 May 1997 (06.05.97)	US
60/046,655	16 May 1997 (16.05.97)	US
60/066,009	14 November 1997 (14.11.97)	US

(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KUNSCH, Charles, A. [US/US]; 4083 Spalding Hollow, Norcross, GA 30092 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). BARASH, Steven, C. [US/US]; 582 College Parkway #303, Rockville, MD 20850 (US).

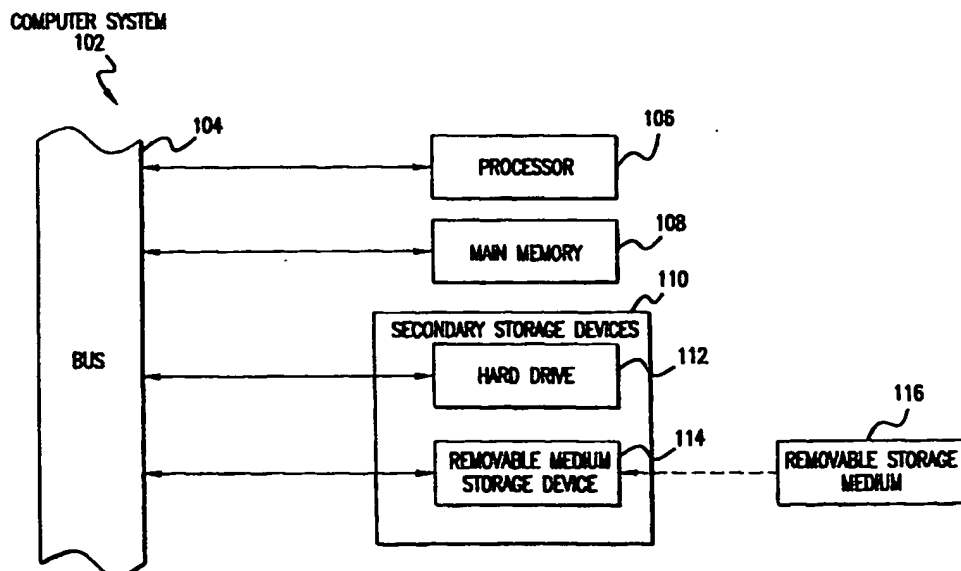
(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: ENTEROCOCCUS FAECALIS POLYNUCLEOTIDES AND POLYPEPTIDES



(57) Abstract

The present invention provides polynucleotide sequences of the genome of *Enterococcus faecalis*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LJ	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						



**Enterococcus faecalis Polynucleotides and Polypeptides**

**FIELD OF THE INVENTION**

5           The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Enterococcus faecalis*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical  
10   development, among others.

**BACKGROUND OF THE INVENTION**

          Enterococci have been recognized as being pathogenic for humans since the turn of the century when they were first described by Thiercelin in 1988 as microscopic  
15   organisms. The genus *Enterococcus* includes the species *Enterococcus faecalis* or *E. faecalis* which is the most common pathogen in the group, accounting for 80 - 90 percent of all enterococcal infections. See Lewis et al. (1990) Eur J. Clin Microbiol Infect Dis. 9:111-117.

          The incidence of enterococcal infections has increased in recent years and  
20   enterococci are now the second most frequently reported nosocomial pathogens. Enterococcal infection is of particular concern because of its resistance to antibiotics. Recent attention has focused on enterococci not only because of their increasing role in nosocomial infections, but also because of their remarkable and increasing resistance to antimicrobial agents. These factors are mutually reinforcing since resistance allows  
25   enterococci to survive in an environment in which antimicrobial agents are heavily used; the hospital setting provides the antibiotics which eliminate or suppress susceptible bacteria, thereby providing a selective advantage for resistant organisms, and the hospital also provides the potential for dissemination of resistant enterococci via the usual routes of hand and environmental contamination.

30           Antimicrobial resistance can be divided into two general types, inherent or intrinsic property and that which is acquired. The genes for intrinsic resistance, like other species characteristics, appear to reside on the chromosome. Acquired resistance results from either a mutation in the existing DNA or acquisition of new DNA. The various inherent traits expressed by enterococci include resistance to semisynthetic  
35   penicillinase-resistant penicillins, cephalosporins, low levels of aminoglycosides, and low levels of clindamycin. Examples of acquired resistance include resistance to

chloramphenicol, erythromycin, high levels of clindamycin, tetracycline, high levels of aminoglycosides, penicillin by means of penicillinase, fluoroquinolones, and vancomycin. Resistance to high levels of penicillin without penicillinase and resistance to fluoroquinolones are not known to be plasmid or transposon mediated and presumably are due to mutation(s).

Although the main reservoir for enterococci in humans is the gastrointestinal tract, the bacteria can also reside in the gallbladder, urethra and vagina.

*E. faecalis* has emerged as an important pathogen in endocarditis, bacteremia, urinary tract infections (UTIs), intraabdominal infections, soft tissue infections, and neonatal sepsis (Lewis 1990, *supra*). In the 1970s and 1980s enterococci became firmly established as major nosocomial pathogens. They are now the fourth leading cause of hospital-acquired infection and the third leading cause of bacteremia in the United States. Fatality ratios for enterococcal bacteremia range from 12% to 68%, with death due to enterococcal sepsis in 4 to 50% of these cases. See Emori, T.G. (1993) Clin. Microbiol. Rev. 6:428-442.

The ability of enterococci to colonize the gastrointestinal tract, plus the many intrinsic and acquired resistance traits, means that these organisms, which usually seem to have relatively low intrinsic virulence, are given an excellent opportunity to become secondary invaders. Since nosocomial isolates of enterococci have displayed resistance to essentially every useful antimicrobial agent, it will likely become increasingly difficult to successfully treat and control enterococcal infections. Particularly when the various resistance genes come together in a single strain, an event almost certain to occur at some time in the future.

The etiology of diseases mediated or exacerbated by *Enterococcus faecalis*, involves the programmed expression of *E. faecalis* genes, and that characterizing these genes and their patterns of expression would dramatically add to our understanding of the organism and its host interactions. Knowledge of the *E. faecalis* gene and genomic organization would improve our understanding of disease etiology and lead to improved and new ways of preventing, treating and diagnosing diseases. Thus, there is a need to characterize the genome of *E. faecalis* and for polynucleotides of this organism.

#### **SUMMARY OF THE INVENTION**

The present invention is based on the sequencing of fragments of the *Enterococcus faecalis* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS:1-982.

The present invention provides the nucleotide sequence of hundreds of contigs of the *Enterococcus faecalis* genome, which are listed in tables below and set out in the

Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-982.

5       The present invention further provides nucleotide sequences which are at least 95%, 96%, 97%, 98%, and 99%, identical to the nucleotide sequences of SEQ ID NOS:1-982.

10       The nucleotide sequence of SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-982 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and  
15       hybrids of these categories such as magnetic/optical storage media.

      The present invention further provides systems, particularly computer-based systems which contain the sequence information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Enterococcus faecalis* genome.

20       Another embodiment of the present invention is directed to fragments of the *Enterococcus faecalis* genome having particular structural or functional attributes. Such fragments of the *Enterococcus faecalis* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs, fragments which modulate the expression of an operably linked ORF,  
25       hereinafter referred to as expression modulating fragments or EMFs, and fragments which can be used to diagnose the presence of *Enterococcus faecalis* in a sample, hereinafter referred to as diagnostic fragments or DFs.

      Each of the ORFs in fragments of the *Enterococcus faecalis* genome disclosed in Tables 1-3, and the EMFs found 5' prime of the initiation codon, can be used in numerous  
30       ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

35       The present invention further includes recombinant constructs comprising one or more fragments of the *Enterococcus faecalis* genome of the present invention. The

recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Enterococcus faecalis* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Enterococcus faecalis* genome of the present invention. The host cells  
5 can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to isolated polypeptides and proteins encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins  
10 of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present  
15 invention from cells which have been altered to express them.

The invention further provides methods of obtaining homologs of the fragments of the *Enterococcus faecalis* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as a probe or as primers, and  
20 techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

25 The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the  
30 antibodies of the present invention, or one or more of the DFs of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain  
35 the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising

one of the antibodies, or one of the DFs of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies or hybridized DFs.

Using the isolated proteins of the present invention, the present invention  
5 further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of:  
(a)contacting an agent with an isolated protein encoded by one of the ORFs of the  
10 present invention; and (b)determining whether the agent binds to said protein.

The present genomic sequences of *Enterococcus faecalis* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Enterococcus faecalis* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to  
15 *Enterococcus faecalis* researchers and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will  
20 provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

25

#### **DESCRIPTION OF THE FIGURES**

**FIGURE 1** is a block diagram of a computer system (102) that can be used to implement computer-based systems of the present invention.

30 **FIGURE 2** is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Enterococcus faecalis* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System  
35 Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of

sequence files. The program Sequis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Enterococcus faecalis* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using Extrseq, a  
5 Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq\_filter to trim portions of the sequences with more than 1% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research ( TIGR ) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly  
10 step is loaded into the database with the lassie program. Identification of open reading frames (ORFs) is accomplished by processing contigs with GeneMark, described in Borodovsky, M. and McIninch, J.D. (1993) *Comput. Chem.*, 17:123133. The ORFs are searched against *E. faecalis* sequences from GenBank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol.* 215:  
15 403-410 (1990)). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

#### **DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS**

20 The present invention is based on the sequencing of fragments of the *Enterococcus faecalis* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS: 1-982. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

25 In addition to the aforementioned *Enterococcus faecalis* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS: 1-982 , or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

30 As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-982" refers to any portion of the SEQ ID NOS: 1-982 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Enterococcus faecalis* open reading frames (ORFs ), expression modulating fragment ( EMFs ) and fragments which can be used to diagnose the presence of *Enterococcus faecalis* in a sample (DFs ). A non-limiting identification  
35 of preferred representative fragments is provided in Tables 1-3. As discussed in detail below, the information provided in SEQ ID NOS:1-982 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the

art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Enterococcus faecalis* proteins.

The present invention is further directed to nucleic acid molecules encoding portions or fragments of the nucleotide sequences described herein. Fragments include portions of the nucleotide sequences of Table 1-3 and SEQ ID NOS:1-982, at least 10 contiguous nucleotides in length selected from any two integers, one of which representing a 5' nucleotide position and a second of which representing a 3' nucleotide position, where the first nucleotide for each nucleotide sequence in SEQ ID NOS:1-982 is position 1. That is, every combination of a 5' and 3' nucleotide position that a fragment at least 10 contiguous nucleotides in length could occupy is included in the invention. At least means a fragment may be 10 contiguous nucleotide bases in length or any integer between 10 and the length of an entire nucleotide sequence of SEQ ID NOS:1-982 minus 1. Therefore, included in the invention are contiguous fragments specified by any 5' and 3' nucleotide base positions of a nucleotide sequences of SEQ ID NOS:1-982 wherein the contiguous fragment is any integer between 10 and the length of an entire nucleotide sequence minus 1.

Further, the invention includes polynucleotides comprising fragments specified by size, in nucleotides, rather than by nucleotide positions. The invention includes any fragment size, in contiguous nucleotides, selected from integers between 10 and the length of an entire nucleotide sequence minus 1. Preferred sizes of contiguous nucleotide fragments include 20 nucleotides, 30 nucleotides, 40 nucleotides, 50 nucleotides. Other preferred sizes of contiguous nucleotide fragments, which may be useful as diagnostic probes and primers, include fragments 50-300 nucleotides in length which include, as discussed above, fragment sizes representing each integer between 50-300. Larger fragments are also useful according to the present invention corresponding to most, if not all, of the nucleotide sequences shown in SEQ ID NOS:1-982. The preferred sizes are, of course, meant to exemplify not limit the present invention as all size fragments, representing any integer between 10 and the length of an entire nucleotide sequence minus 1, of each SEQ ID NO., are included in the invention.

The present invention also provides for the exclusion of any fragment, specified by 5' and 3' base positions or by size in nucleotide bases as described above for any nucleotide sequence of SEQ ID NOS:1-982. Any number of fragments of nucleotide sequences in SEQ ID NOS:1-982, specified by 5' and 3' base positions or by size in nucleotides, as described above, may be excluded from the present invention.

While the presently disclosed sequences of SEQ ID NOS:1-982 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a nucleotide

sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-982.

However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-982 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS: 1-982 will be well within the skill of the art. The present disclosure

5 makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotides may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied  
10 Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS: 1-982 were  
15 corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-982.

As discussed elsewhere herein, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning  
20 and sequencing DNA. A wide variety of *Enterococcus faecalis* strains that can be used to prepare *E. faecalis* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC). While the present invention is enabled by the sequences and other information herein disclosed, the *E. faecalis* strain  
25 that provided the DNA of the present Sequence Listing, Strain V586, kindly provided by Dr. Michael Gilmore, University of Oklahoma, has been deposited in the ATCC, as a convenience to those of skill in the art. The *E. faecalis* strain V586 was deposited 2 May 1997 at the ATCC, 10801 University Blvd. Manassas, VA 20110-2209, and given accession number 55969. The provision of the deposits is not a waiver of any rights of  
30 the inventors or their assignees in the present subject matter.

The nucleotide sequences of the genomes from different strains of *Enterococcus faecalis* differ somewhat. However, the nucleotide sequences of the genomes of all *Enterococcus faecalis* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS: 1-982. Nearly all will be at least 99%  
35 identical and the great majority will be 99.9% identical.

The present application is further directed to nucleic acid molecules at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleic acid sequence shown in SEQ ID NOS:



1-982. The above nucleic acid sequences are included irrespective of whether they encode a polypeptide having *E. faecalis* activity. This is because even where a particular nucleic acid molecule does not encode a polypeptide having *E. faecalis* activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having *E. faecalis* activity include, *inter alia*, isolating an *E. faecalis* gene or allelic variants thereof from a DNA library, and detecting *E. faecalis* mRNA expression samples, environmental samples, suspected of containing *E. faecalis* by Northern Blot analysis.

Preferred, are nucleic acid molecules having sequences at least 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in SEQ ID NOS: 1-982, which do, in fact, encode a polypeptide having *E. faecalis* protein activity. By "a polypeptide having *E. faecalis* activity" is intended polypeptides exhibiting activity similar, but not necessarily identical, to an activity of the *E. faecalis* protein of the invention, as measured in a particular biological assay suitable for measuring activity of the specified protein.

Due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the nucleic acid molecules having a sequence at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the nucleic acid sequences shown in SEQ ID NOS: 1-982 will encode a polypeptide having *E. faecalis* protein activity. In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having *E. faecalis* protein activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid), as further described below.

The biological activity or function of the polypeptides of the present invention are expected to be similar or identical to polypeptides from other bacteria that share a high degree of structural identity/similarity. Tables 1 and 2 lists accession numbers and descriptions for the closest matching sequences of polypeptides available through Genbank. It is therefore expected that the biological activity or function of the polypeptides of the present invention will be similar or identical to those polypeptides from other bacterial genuses, species, or strains listed in Tables 1 and 2.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that

the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the *E. faecalis* polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted, inserted, or substituted with another nucleotide. The query sequence may be an entire sequence shown in SEQ ID NOS: 1-982, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. See Brutlag et al. (1990) Comp. App. Biosci. 6:237-245. In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by first converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only nucleotides outside the 5' and 3' nucleotides of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 nucleotide subject sequence is aligned to a 100 nucleotide query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 nucleotides at 5' end. The 10 unpaired nucleotides represent 10% of the sequence (number of nucleotides at the 5' and 3' ends not matched/total number of nucleotides in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 nucleotides were perfectly matched the final percent identity would be 90%. In another example, a 90 nucleotide subject sequence is compared with a 100 nucleotide query sequence. This time the deletions are internal deletions so that there are no nucleotides on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only nucleotides 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

#### COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-982 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, provided refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; i.e., a nucleotide sequence provided in SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-982. Such a manufacture provides a large portion of the *Enterococcus faecalis* genome and parts thereof (e.g., a *Enterococcus faecalis* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Enterococcus faecalis* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily

appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently know methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention. A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially- available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS: 1-982 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Enterococcus faecalis* genome which contain homology to ORFs or proteins from both *Enterococcus faecalis* and from other organisms. Among the ORFs discussed herein are protein encoding fragments of the *Enterococcus faecalis* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful

metabolites, proteins to be used as vaccines or in the generation of immuno-therapeutic reagents, or as drug screening targets.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are  
5 designed to identify, among other things, commercially important fragments of the *Enterococcus faecalis* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the  
10 present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention  
15 and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the  
20 present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences  
25 which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI). A skilled artisan can readily recognize that any one of the  
30 available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a  
35 random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important

fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Enterococcus faecalis* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Enterococcus faecalis* genome. In the present examples, implementing software which implement the BLAST algorithm, described in Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410, is used to identify open reading frames within the *Enterococcus faecalis* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing and processing the genomic sequence (such as search tools, comparing tools, *etc.*) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

### BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to isolated fragments of the *Enterococcus faecalis* genome. The fragments of the *Enterococcus faecalis* genome of the present invention include, but are not limited to fragments which encode peptides, hereinafter open reading frames (ORFs), fragments which modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Enterococcus faecalis* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Enterococcus faecalis* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-982, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Enterococcus faecalis* DNA can be enzymatically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate a *Enterococcus faecalis* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS:1-982. Well known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library or *Enterococcus faecalis* genomic DNA. Thus, given the availability of SEQ ID NOS:1-982, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-982 using methods set out above, those of skill will be enabled by the present

disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA. As used  
5 herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein. Each sequence of SEQ ID NOS:1-982, however, begins and ends with a termination codon. For purposes of numbering and reference to polynucleotide and polypeptide sequences the entire sequence of each sequence of SEQ ID NOS:1-982 is included with the first  
10 nucleotide being position 1. Therefore, for reference purposes the numbering used in the present invention is that provided in the sequence listing for SEQ ID NOS:1-982.

Tables 1, 2, and 3 list ORFs in the *Enterococcus faecalis* genomic contigs of the present invention that were identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It  
15 will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive, or more selective lists.

Table 1 sets out ORFs in the *Enterococcus faecalis* contigs of the present invention that over a continuous region of at least 50 bases are 95% or more identical (by  
20 BLAST analysis) to a nucleotide sequence available through GenBank in March, 1997.

Table 2 sets out ORFs in the *Enterococcus faecalis* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through GenBank in March, 1997.

Table 3 sets out ORFs in the *Enterococcus faecalis* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence  
25 available through GenBank in March, 1997.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the coordinate of the first nucleotide of the ORF, counting from the 5' end of the contig  
30 strand; the fourth column indicates the coordinate of the final nucleotide of the ORF, counting from the 5' end of the contig strand.

In Tables 1 and 2, column five lists the Reference for the closest matching sequence available through GenBank. These reference numbers are the database entry numbers commonly used by those of skill in the art, who will be familiar with their  
35 denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column six in Tables 1 and 2 provides the gene name of the matching sequence.



In Table 1, column seven provides the nucleotide BLAST percent identity score from the comparison of the ORF and the GenBank sequence, column eight indicates the length in nucleotides of the highest scoring segment pair identified by the BLAST identity analysis, and column nine provides the total length of the ORF in nucleotides.

5 In Table 2, column seven provides the protein BLAST percent similarity of the highest scoring segment pair identified, column eight provides the percent identity of the highest scoring segment pair, and column nine provides the total length of the ORF in nucleotides.

The concepts of percent identity and percent similarity of two polypeptide  
10 sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (*e.g.*, at positions 1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.*, possessed similar biochemical characteristics).  
15 Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list percent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the percent identity of the highest scoring segment pair in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in  
20 the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Enterococcus faecalis* genome other than those listed in  
25 Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

30 As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression or an operably linked ORF in response to a specific regulatory factor or  
35 physiological event.

EMF sequences can be identified within the contigs of the *Enterococcus faecalis* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment,

or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the

5 *Enterococcus faecalis* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector.

10 An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker

15 sequence. A more detailed discussion of various marker sequences is provided below.

A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate

20 conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Enterococcus faecalis* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Enterococcus*

25 *faecalis* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to

30 the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 99% and preferably 99.9% identical to SEQ ID NOS:1-982, with a sequence from another isolate of the same species. Furthermore, to accommodate

35 codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the

coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.*, sequence both  
5 strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Enterococcus faecalis* origin isolated by using part or all of the fragments in question as a probe or primer.

Each of the ORFs of the *Enterococcus faecalis* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5 to the ORFs, can be used as polynucleotide reagents in  
10 numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particularly *Enterococcus faecalis*. Especially preferred in this regard are ORFs such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Enterococcus faecalis*. Also  
15 particularly preferred are ORFs that can be used to distinguish between strains of *Enterococcus faecalis*, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA,  
20 both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for  
25 use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for  
30 example, Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* 56:560 (1991) and *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)).

35 The present invention further provides recombinant constructs comprising one or more fragments of the *Enterococcus faecalis* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention

comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Enterococcus faecalis* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably  
5 linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the  
10 present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBS SK (+ or -), pBS KS (+ or -), pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (available from Stratagene) pSVK3, pBPV, pMSG,  
15 pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV  
20 immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the *Enterococcus faecalis* genomic fragments and contigs of the  
25 present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct  
30 comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

35 A host cell containing one of the fragments of the *Enterococcus faecalis* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can

be used to produce a heterologous protein under the control of the EMF. The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide  
5 fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs depicted in Tables 2 and 3 which encode proteins.

10 A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are  
15 useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and  
20 purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography.

25 The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. Preferred polypeptides and proteins of the present invention are polypeptides and proteins coded for by the polynucleotides of SEQ ID NOS:1-982, wherein the polypeptides and proteins are coded in the same frame as the termination codon at the end of each sequence of SEQ  
30 ID NOS:1-982. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells  
35 in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of the *E. faecalis* polypeptide can be substantially purified by the one-step method described by Smith et al. (1988) Gene 67:31-40. Polypeptides of the invention also can be purified  
5 from natural or recombinant sources using antibodies directed against the polypeptides of the invention in methods which are well known in the art of protein purification.

The invention further provides for isolated *E. faecalis* polypeptides comprising an amino acid sequence selected from the group including: (a) the amino acid sequence of a full-length *E. faecalis* polypeptide having the complete amino acid sequence from the  
10 first methionine codon to the termination codon of each sequence listed in SEQ ID NOS:1-982, wherein said termination codon is at the end of each SEQ ID NO: and said first methionine is the first methionine in frame with said termination codon; and (b) the amino acid sequence of a full-length *E. faecalis* polypeptide having the complete amino acid sequence in (a) excepting the N-terminal methionine.

15 The polypeptides of the present invention also include polypeptides having an amino acid sequence at least 80% identical, more preferably at least 90% identical, and still more preferably 95%, 96%, 97%, 98% or 99% identical to those described in (a) and (b) above.

The present invention is further directed to polynucleotide encoding portions or  
20 fragments of the amino acid sequences described herein as well as to portions or fragments of the isolated amino acid sequences described herein. Fragments include portions of the amino acid sequences described herein, are at least 5 contiguous amino acid in length, are selected from any two integers, one of which representing a N-terminal position. The initiation codon of the polypeptides of the present inventions position 1. The initiation  
25 codon (position 1) for purposes of the present invention is the first methionine codon of each sequence of SEQ ID NOS:1-982 which is in frame with the termination codon at the end of each said sequence. Every combination of a N-terminal and C-terminal position that a fragment at least 5 contiguous amino acid residues in length could occupy, on any given amino acid sequence encoded by a sequence of SEQ ID NOS:1-982 is included in the  
30 invention, i.e., from initiation codon up to the termination codon. At least means a fragment may be 5 contiguous amino acid residues in length or any integer between 5 and the number of residues in a full length amino acid sequence minus 1. Therefore, included in the invention are contiguous fragments specified by any N-terminal and C-terminal positions of amino acid sequence set forth in SEQ ID NOS:1-982 wherein the contiguous  
35 fragment is any integer between 5 and the number of residues in a full length sequence minus 1.

Further, the invention includes polypeptides comprising fragments specified by size, in amino acid residues, rather than by N-terminal and C-terminal positions. The invention includes any fragment size, in contiguous amino acid residues, selected from integers between 5 and the number of residues in a full length sequence minus 1. Preferred sizes of contiguous polypeptide fragments include about 5 amino acid residues, about 10 amino acid residues, about 20 amino acid residues, about 30 amino acid residues, about 40 amino acid residues, about 50 amino acid residues, about 100 amino acid residues, about 200 amino acid residues, about 300 amino acid residues, and about 400 amino acid residues. The preferred sizes are, of course, meant to exemplify, not limit, the present invention as all size fragments representing any integer between 5 and the number of residues in a full length sequence minus 1 are included in the invention. The present invention also provides for the exclusion of any fragments specified by N-terminal and C-terminal positions or by size in amino acid residues as described above. Any number of fragments specified by N-terminal and C-terminal positions or by size in amino acid residues as described above may be excluded.

The above fragments need not be active since they would be useful, for example, in immunoassays, in epitope mapping, epitope tagging, to generate antibodies to a particular portion of the protein, as vaccines, and as molecular weight markers.

Further polypeptides of the present invention include polypeptides which have at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98% or 99% similarity to those described above.

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of a *E. faecalis* polypeptide having an amino acid sequence which contains at least one conservative amino acid substitution, but not more than 50 conservative amino acid substitutions, not more than 40 conservative amino acid substitutions, not more than 30 conservative amino acid substitutions, and not more than 20 conservative amino acid substitutions. Also provided are polypeptides which comprise the amino acid sequence of a *E. faecalis* polypeptide, having at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 conservative amino acid substitutions.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the

reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

- 5           As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to the amino acid sequences encoded by the sequences of SEQ ID NOS:1-982, as described hererin, can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence,  
10   also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al., (1990) Comp. App. Biosci. 6:237-245. In a sequence alignment the query and subject sequences are both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2,  
15   Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

- If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, the results, in percent identity, must be  
20   manually corrected. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a  
25   corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the  
30   present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query amino acid residues outside the farthest N- and C-terminal residues of the subject sequence.

- For example, a 90 amino acid residue subject sequence is aligned with a 100 residue  
35   query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not match/align with the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the



sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject  
5 sequence is compared with a 100 residue query sequence. This time the deletions are internal so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which  
10 are not matched/aligned with the query sequence are manually corrected. No other manual corrections are to be made for the purposes of the present invention.

The above polypeptide sequences are included irrespective of whether they have their normal biological activity. This is because even where a particular polypeptide molecule does not have biological activity, one of skill in the art would still know how to  
15 use the polypeptide, for instance, as a vaccine or to generate antibodies. Other uses of the polypeptides of the present invention that do not have *E. faecalis* activity include, *inter alia*, as epitope tags, in epitope mapping, and as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods known to those of skill in the art.

20 As described below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting *E. faecalis* protein expression or as agonists and antagonists capable of enhancing or inhibiting *E. faecalis* protein function. Further, such polypeptides can be used in the yeast two-hybrid system to "capture" *E. faecalis* protein binding proteins which are also  
25 candidate agonists and antagonists according to the present invention. *See, e.g.*, Fields et al. (1989) Nature 340:245-246.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B.*  
30 *subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast)  
35 expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*,

*E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides.

Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Enterococcus faecalis* genome and short  
5 oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

Recombinant expression vehicle or "vector" refers to a plasmid or phage or virus  
10 or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancer  
15 and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell.  
20 Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a  
25 recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

30 Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring  
35 Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas* and *Streptomyces*. Others may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available from Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5 flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Enterococcus faecalis*, of the fragments of the *Enterococcus faecalis* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Enterococcus faecalis* is defined as a homolog of a fragment of the *Enterococcus faecalis* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Enterococcus faecalis* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the

sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among those are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-982 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ ID NOS:1-982 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis *et al.*, *PCR Protocols*, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-982 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS: 1-982, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C in 6X SSPE and 50% formamide, and washing at 50- 65°C in 0.5X SSPE) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPE and 40-45% formamide, and washing at 42°C in 0.5X SSPE), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-982, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-982, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, hybridizing at 50- 65°C in 5X SSPE and 50% formamide, and washing at 50- 65°C in 0.5X SSPE), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPE and 40-45% formamide, and washing at 42°C in 0.5X SSPE), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Enterococcus faecalis*.

5

## ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the polypeptide. Such identifications permit one skilled in the art to use the *Enterococcus faecalis* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., MacMillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESSES, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

20

### 1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

30

The various metabolic pathways present in *Enterococcus faecalis* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-982.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

35

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*, *Symbiosis* 21:79 (1986) and Voragen *et al.* in *Biocatalysts In Agricultural Biotechnology*, Whitaker *et al.*, Eds., *American Chemical Society Symposium Series* 389:93 (1989) .

The metabolism of sugars is an important aspect of the primary metabolism of *Enterococcus faecalis*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters* 1:21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in *GENE MANIPULATIONS AND FUNGI*; Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta.* 872:83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose- produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, *Starch* 40:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See  
5 Faultman *et al.*, Acid Proteases Structure Function and Biology, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, Industrial Enzymes, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, Report Industrial Enzymes by 1990, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the  
10 microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London* 310:227 (1985) and Poserke, *Journal of the American Oil Chemist Society* 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to  
15 facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those  
20 scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions,  
25 trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary  
30 to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other hand, has been described in detail by Bud *et al.*, *Chemistry in Britain* (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of  
35 amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo- selective synthesis of only L-amino acids and generally possess uniformly high



catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination.

5

## 2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be  
10 used to generate an antibody which selectively binds the protein.

*E. faecalis* protein-specific antibodies for use in the present invention can be raised against the intact *E. faecalis* protein or an antigenic polypeptide fragment thereof, which may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids),  
15 without a carrier.

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules, single chain whole antibodies, and antibody fragments. Antibody fragments of the present invention include Fab and F(ab')<sub>2</sub> and other fragments including single-chain Fvs (scFv) and disulfide-linked Fvs (sdFv). Also included in the  
20 present invention are chimeric and humanized monoclonal antibodies and polyclonal antibodies specific for the polypeptides of the present invention. The antibodies of the present invention may be prepared by any of a variety of methods. For example, cells expressing a polypeptide of the present invention or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing  
25 polyclonal antibodies. For example, a preparation of *E. faecalis* polypeptide or fragment thereof is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In a preferred method, the antibodies of the present invention are monoclonal  
30 antibodies or binding fragments thereof. Such monoclonal antibodies can be prepared using hybridoma technology. See, e.g., Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988); Hammerling, et al., in: MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS 563-681 (Elsevier, N.Y., 1981). Fab and F(ab')<sub>2</sub> fragments may be produced by proteolytic cleavage, using  
35 enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')<sub>2</sub> fragments). Alternatively, *E. faecalis* polypeptide-binding fragments, chimeric, and humanized antibodies can be produced through the application of recombinant DNA

technology or through synthetic chemistry using methods known in the art.

Alternatively, additional antibodies capable of binding to the polypeptide antigen of the present invention may be produced in a two-step procedure through the use of anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and that, therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, *E. faecalis* polypeptide-specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the *E. faecalis* polypeptide-specific antibody can be blocked by the *E. faecalis* polypeptide antigen. Such antibodies comprise anti-idiotypic antibodies to the *E. faecalis* polypeptide-specific antibody and can be used to immunize an animal to induce formation of further *E. faecalis* polypeptide-specific antibodies.

Antibodies and fragments thereof of the present invention may be described by the portion of a polypeptide of the present invention recognized or specifically bound by the antibody. Antibody binding fragments of a polypeptide of the present invention may be described or specified in the same manner as for polypeptide fragments discussed above, i.e., by N-terminal and C-terminal positions or by size in contiguous amino acid residues. Any number of antibody binding fragments, of a polypeptide of the present invention, specified by N-terminal and C-terminal positions or by size in amino acid residues, as described above, may also be excluded from the present invention. Therefore, the present invention includes antibodies that specifically bind a particularly described fragment of a polypeptide of the present invention and allows for the exclusion of the same.

Antibodies and fragments thereof of the present invention may also be described or specified in terms of their cross-reactivity. Antibodies and fragments that do not bind polypeptides of any other species of *Enterococcus* other than *E. faecalis* are included in the present invention. Likewise, antibodies and fragments that bind only species of *Enterococcus*, i.e. antibodies and fragments that do not bind bacteria from any genus other than *Enterococcus*, are included in the present invention.

### 3. Diagnostic and Detection Assays and Kits

The present invention further relates to methods for assaying enterococcal infection in an animal by detecting the expression of genes encoding enterococcal polypeptides of the present invention. The methods comprise analyzing tissue or body fluid from the animal for *Enterococcus*-specific antibodies, nucleic acids, or proteins. Analysis of nucleic acid specific to *Enterococcus* is assayed by PCR or hybridization

techniques using nucleic acid sequences of the present invention as either hybridization probes or primers. See, e.g., Sambrook et al. Molecular cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 2nd ed., 1989, page 54 reference); Ereemeeva et al. (1994) J. Clin. Microbiol. 32:803-810 (describing differentiation among spotted fever  
5 group *Rickettsiae* species by analysis of restriction fragment length polymorphism of PCR-amplified DNA) and Chen et al. 1994 J. Clin. Microbiol. 32:589-595 (detecting *B. burgdorferi* nucleic acids via PCR).

Where diagnosis of a disease state related to infection with *Enterococcus* has already been made, the present invention is useful for monitoring progression or  
10 regression of the disease state whereby patients exhibiting enhanced *Enterococcus* gene expression will experience a worse clinical outcome relative to patients expressing these gene(s) at a lower level.

By "biological sample" is intended any biological sample obtained from an animal, cell line, tissue culture, or other source which contains *Enterococcus* polypeptide, mRNA,  
15 or DNA. Biological samples include body fluids (such as saliva, blood, plasma, urine, mucus, synovial fluid, etc.) tissues (such as muscle, skin, and cartilage) and any other biological source suspected of containing *Enterococcus* polypeptides or nucleic acids. Methods for obtaining biological samples such as tissue are well known in the art.

The present invention is useful for detecting diseases related to *Enterococcus*  
20 infections in animals. Preferred animals include monkeys, apes, cats, dogs, birds, cows, pigs, mice, horses, rabbits and humans. Particularly preferred are humans.

Total RNA can be isolated from a biological sample using any suitable technique such as the single-step guanidinium-thiocyanate-phenol-chloroform method described in Chomczynski et al. (1987) Anal. Biochem. 162:156-159. mRNA encoding *Enterococcus*  
25 polypeptides having sufficient homology to the nucleic acid sequences identified in SEQ ID NOS:1-982 to allow for hybridization between complementary sequences are then assayed using any appropriate method. These include Northern blot analysis, S1 nuclease mapping, the polymerase chain reaction (PCR), reverse transcription in combination with the polymerase chain reaction (RT-PCR), and reverse transcription in combination  
30 with the ligase chain reaction (RT-LCR).

Northern blot analysis can be performed as described in Harada et al. (1990) Cell 63:303-312. Briefly, total RNA is prepared from a biological sample as described above. For the Northern blot, the RNA is denatured in an appropriate buffer (such as glyoxal/dimethyl sulfoxide/sodium phosphate buffer), subjected to agarose gel  
35 electrophoresis, and transferred onto a nitrocellulose filter. After the RNAs have been linked to the filter by a UV linker, the filter is prehybridized in a solution containing formamide, SSC, Denhardt's solution, denatured salmon sperm, SDS, and sodium

phosphate buffer. A *E. faecalis* polynucleotide sequence shown in SEQ ID NOS:1-982 labeled according to any appropriate method (such as the <sup>32</sup>P-multiprimered DNA labeling system (Amersham)) is used as probe. After hybridization overnight, the filter is washed and exposed to x-ray film. DNA for use as probe according to the present invention is

5 described in the sections above and will preferably at least 15 nucleotides in length.

S1 mapping can be performed as described in Fujita et al. (1987) Cell 49:357-367. To prepare probe DNA for use in S1 mapping, the sense strand of an above-described *E. faecalis* DNA sequence of the present invention is used as a template to synthesize labeled antisense DNA. The antisense DNA can then be digested using an appropriate restriction

10 endonuclease to generate further DNA probes of a desired length. Such antisense probes are useful for visualizing protected bands corresponding to the target mRNA (i.e., mRNA encoding *Enterococcus* polypeptides).

Levels of mRNA encoding *Enterococcus* polypeptides are assayed, for e.g., using the RT-PCR method described in Makino et al. (1990) Technique 2:295-301. By this

15 method, the radioactivities of the "amplicons" in the polyacrylamide gel bands are linearly related to the initial concentration of the target mRNA. Briefly, this method involves adding total RNA isolated from a biological sample in a reaction mixture containing a RT primer and appropriate buffer. After incubating for primer annealing, the mixture can be supplemented with a RT buffer, dNTPs, DTT, RNase inhibitor and

20 reverse transcriptase. After incubation to achieve reverse transcription of the RNA, the RT products are then subject to PCR using labeled primers. Alternatively, rather than labeling the primers, a labeled dNTP can be included in the PCR reaction mixture. PCR amplification can be performed in a DNA thermal cycler according to conventional techniques. After a suitable number of rounds to achieve amplification, the PCR reaction

25 mixture is electrophoresed on a polyacrylamide gel. After drying the gel, the radioactivity of the appropriate bands (corresponding to the mRNA encoding the *Enterococcus* polypeptides of the present invention) are quantified using an imaging analyzer. RT and PCR reaction ingredients and conditions, reagent and gel concentrations, and labeling methods are well known in the art. Variations on the

30 RT-PCR method will be apparent to the skilled artisan. Other PCR methods that can detect the nucleic acid of the present invention can be found in PCR PRIMER: A LABORATORY MANUAL (C.W. Dieffenbach et al. eds., Cold Spring Harbor Lab Press, 1995).

The polynucleotides of the present invention, including both DNA and RNA, may

35 be used to detect polynucleotides of the present invention or Enterococcal species including *E. faecalis* using bio chip technology. The present invention includes both high density chip arrays (>1000 oligonucleotides per cm<sup>2</sup>) and low density chip arrays (<1000

oligonucleotides per cm<sup>2</sup>). Bio chips comprising arrays of polynucleotides of the present invention may be used to detect Enterococcal species, including *E. faecalis*, in biological and environmental samples and to diagnose an animal, including humans, with an *E. faecalis* or other Enterococcal infection. The bio chips of the present invention may  
5 comprise polynucleotide sequences of other pathogens including bacteria, viral, parasitic, and fungal polynucleotide sequences, in addition to the polynucleotide sequences of the present invention, for use in rapid differential pathogenic detection and diagnosis. The bio chips can also be used to monitor an *E. faecalis* or other Enterococcal infections and to monitor the genetic changes (deletions, insertions, mismatches, etc.) in response to  
10 drug therapy in the clinic and drug development in the laboratory. The bio chip technology comprising arrays of polynucleotides of the present invention may also be used to simultaneously monitor the expression of a multiplicity of genes, including those of the present invention. The polynucleotides used to comprise a selected array may be specified in the same manner as for the fragments, i.e, by their 5' and 3' positions or  
15 length in contiguous base pairs and include from. Methods and particular uses of the polynucleotides of the present invention to detect Enterococcal species, including *E. faecalis*, using bio chip technology include those known in the art and those of: U.S. Patent Nos. 5510270, 5545531, 5445934, 5677195, 5532128, 5556752, 5527681, 5451683, 5424186, 5607646, 5658732 and World Patent Nos. WO/9710365,  
20 WO/9511995, WO/9743447, WO/9535505, each incorporated herein in their entireties.

Biosensors using the polynucleotides of the present invention may also be used to detect, diagnose, and monitor *E. faecalis* or other Enterococcal species and infections thereof. Biosensors using the polynucleotides of the present invention may also be used to detect particular polynucleotides of the present invention. Biosensors using the  
25 polynucleotides of the present invention may also be used to monitor the genetic changes (deletions, insertions, mismatches, etc.) in response to drug therapy in the clinic and drug development in the laboratory. Methods and particular uses of the polynucleotides of the present invention to detect Enterococcal species, including *E. faecalis*, using biosensors include those known in the art and those of: U.S. Patent Nos 5721102, 5658732,  
30 5631170, and World Patent Nos. WO97/35011, WO/9720203, each incorporated herein in their entireties.

Thus, the present invention includes both bio chips and biosensors comprising polynucleotides of the present invention and methods of their use.

Assaying *Enterococcus* polypeptide levels in a biological sample can occur using  
35 any art-known method, such as antibody-based techniques. For example, *Enterococcus* polypeptide expression in tissues can be studied with classical immunohistological methods. In these, the specific recognition is provided by the primary antibody

(polyclonal or monoclonal) but the secondary detection system can utilize fluorescent, enzyme, or other conjugated secondary antibodies. As a result, an immunohistological staining of tissue section for pathological examination is obtained. Tissues can also be extracted, e.g., with urea and neutral detergent, for the liberation of *Enterococcus* polypeptides for Western-blot or dot/slot assay. See, e.g., Jalkanen, M. et al. (1985) J. Cell. Biol. 101:976-985; Jalkanen, M. et al. (1987) J. Cell. Biol. 105:3087-3096. In this technique, which is based on the use of cationic solid phases, quantitation of a *Enterococcus* polypeptide can be accomplished using an isolated *Enterococcus* polypeptide as a standard. This technique can also be applied to body fluids.

Other antibody-based methods useful for detecting *Enterococcus* polypeptide gene expression include immunoassays, such as the ELISA and the radioimmunoassay (RIA). For example, a *Enterococcus* polypeptide-specific monoclonal antibodies can be used both as an immunoabsorbent and as an enzyme-labeled probe to detect and quantify a *Enterococcus* polypeptide. The amount of a *Enterococcus* polypeptide present in the sample can be calculated by reference to the amount present in a standard preparation using a linear regression computer algorithm. Such an ELISA is described in Iacobelli et al. (1988) Breast Cancer Research and Treatment 11:19-30. In another ELISA assay, two distinct specific monoclonal antibodies can be used to detect *Enterococcus* polypeptides in a body fluid. In this assay, one of the antibodies is used as the immunoabsorbent and the other as the enzyme-labeled probe.

The above techniques may be conducted essentially as a "one-step" or "two-step" assay. The "one-step" assay involves contacting the *Enterococcus* polypeptide with immobilized antibody and, without washing, contacting the mixture with the labeled antibody. The "two-step" assay involves washing before contacting the mixture with the labeled antibody. Other conventional methods may also be employed as suitable. It is usually desirable to immobilize one component of the assay system on a support, thereby allowing other components of the system to be brought into contact with the component and readily removed from the sample. Variations of the above and other immunological methods included in the present invention can also be found in Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988).

Suitable enzyme labels include, for example, those from the oxidase group, which catalyze the production of hydrogen peroxide by reacting with substrate. Glucose oxidase is particularly preferred as it has good stability and its substrate (glucose) is readily available. Activity of an oxidase label may be assayed by measuring the concentration of hydrogen peroxide formed by the enzyme-labeled antibody/substrate reaction. Besides enzymes, other suitable labels include radioisotopes, such as iodine ( $^{125}\text{I}$ ,  $^{121}\text{I}$ ), carbon

( $^{14}\text{C}$ ), sulphur ( $^{35}\text{S}$ ), tritium ( $^3\text{H}$ ), indium ( $^{112}\text{In}$ ), and technetium ( $^{99\text{m}}\text{Tc}$ ), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

Further suitable labels for the *Enterococcus* polypeptide-specific antibodies of the present invention are provided below. Examples of suitable enzyme labels include malate dehydrogenase, Enterococcal nuclease, delta-5-steroid isomerase, yeast-alcohol dehydrogenase, alpha-glycerol phosphate dehydrogenase, triose phosphate isomerase, peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase, and acetylcholine esterase.

Examples of suitable radioisotopic labels include  $^3\text{H}$ ,  $^{111}\text{In}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ ,  $^{51}\text{Cr}$ ,  $^{57}\text{Co}$ ,  $^{58}\text{Co}$ ,  $^{59}\text{Fe}$ ,  $^{75}\text{Se}$ ,  $^{152}\text{Eu}$ ,  $^{90}\text{Y}$ ,  $^{67}\text{Cu}$ ,  $^{217}\text{Bi}$ ,  $^{211}\text{At}$ ,  $^{212}\text{Pb}$ ,  $^{47}\text{Sc}$ ,  $^{109}\text{Pd}$ , etc.  $^{111}\text{In}$  is a preferred isotope where *in vivo* imaging is used since it avoids the problem of dehalogenation of the  $^{125}\text{I}$  or  $^{131}\text{I}$ -labeled monoclonal antibody by the liver. In addition, this radionuclide has a more favorable gamma emission energy for imaging. See, e.g., Perkins et al. (1985) Eur. J. Nucl. Med. 10:296-301; Carasquillo et al. (1987) J. Nucl. Med. 28:281-287. For example,  $^{111}\text{In}$  coupled to monoclonal antibodies with 1-(P-isothiocyanatobenzyl)-DPTA has shown little uptake in non-tumors tissues, particularly the liver, and therefore enhances specificity of tumor localization. See, Esteban et al. (1987) J. Nucl. Med. 28:861-870.

Examples of suitable non-radioactive isotopic labels include  $^{157}\text{Gd}$ ,  $^{55}\text{Mn}$ ,  $^{162}\text{Dy}$ ,  $^{52}\text{Tr}$ , and  $^{56}\text{Fe}$ .

Examples of suitable fluorescent labels include an  $^{152}\text{Eu}$  label, a fluorescein label, an isothiocyanate label, a rhodamine label, a phycoerythrin label, a phycocyanin label, an allophycocyanin label, an o-phthaldehyde label, and a fluorescamine label.

Examples of suitable toxin labels include, *Pseudomonas* toxin, diphtheria toxin, ricin, and cholera toxin.

Examples of chemiluminescent labels include a luminal label, an isoluminal label, an aromatic acridinium ester label, an imidazole label, an acridinium salt label, an oxalate ester label, a luciferin label, a luciferase label, and an aequorin label.

Examples of nuclear magnetic resonance contrasting agents include heavy metal nuclei such as Gd, Mn, and iron.

Typical techniques for binding the above-described labels to antibodies are provided by Kennedy et al. (1976) Clin. Chim. Acta 70:1-31, and Schurs et al. (1977) Clin. Chim. Acta 81:1-40. Coupling techniques mentioned in the latter are the glutaraldehyde method, the periodate method, the dimaleimide method, the m-maleimidobenzyl-N-hydroxy-succinimide ester method, all of which methods are incorporated by reference herein.

In a related aspect, the invention includes a diagnostic kit for use in screening serum containing antibodies specific against *E. faecalis* infection. Such a kit may include an isolated *E. faecalis* antigen comprising an epitope which is specifically immunoreactive with at least one anti-*E. faecalis* antibody. Such a kit also includes  
5 means for detecting the binding of said antibody to the antigen. In specific embodiments, the kit may include a recombinantly produced or chemically synthesized peptide or polypeptide antigen. The peptide or polypeptide antigen may be attached to a solid support.

In a more specific embodiment, the detecting means of the above-described kit  
10 includes a solid support to which said peptide or polypeptide antigen is attached. Such a kit may also include a non-attached reporter-labeled anti-human antibody. In this embodiment, binding of the antibody to the *E. faecalis* antigen can be detected by binding of the reporter labeled antibody to the anti-*E. faecalis* polypeptide antibody.

In a related aspect, the invention includes a method of detecting *E. faecalis*  
15 infection in a subject. This detection method includes reacting a body fluid, preferably serum, from the subject with an isolated *E. faecalis* antigen, and examining the antigen for the presence of bound antibody. In a specific embodiment, the method includes a polypeptide antigen attached to a solid support, and serum is reacted with the support. Subsequently, the support is reacted with a reporter-labeled anti-human antibody. The  
20 support is then examined for the presence of reporter-labeled antibody.

The solid surface reagent employed in the above assays and kits is prepared by known techniques for attaching protein material to solid support material, such as polymeric beads, dip sticks, 96-well plates or filter material. These attachment methods generally include non-specific adsorption of the protein to the support or covalent  
25 attachment of the protein, typically through a free amine group, to a chemically reactive group on the solid support, such as an activated carboxyl, hydroxyl, or aldehyde group. Alternatively, streptavidin coated plates can be used in conjunction with biotinylated antigen(s).

The polypeptides and antibodies of the present invention, including fragments  
30 thereof, may be used to detect Enterococcal species including *E. faecalis* using bio chip and biosensor technology. Bio chip and biosensors of the present invention may comprise the polypeptides of the present invention to detect antibodies, which specifically recognize Enterococcal species, including *E. faecalis*. Bio chip and biosensors of the present invention may also comprise antibodies which specifically recognize the  
35 polypeptides of the present invention to detect Enterococcal species, including *E. faecalis* or specific polypeptides of the present invention. Bio chips or biosensors comprising polypeptides or antibodies of the present invention may be used to detect



Enterococcal species, including *E. faecalis*, in biological and environmental samples and to diagnose an animal, including humans, with an *E. faecalis* or other Enterococcal infection. Thus, the present invention includes both bio chips and biosensors comprising polypeptides or antibodies of the present invention and methods of their use.

5       The bio chips of the present invention may further comprise polypeptide sequences of other pathogens including bacteria, viral, parasitic, and fungal polypeptide sequences, in addition to the polypeptide sequences of the present invention, for use in rapid differential pathogenic detection and diagnosis. The bio chips of the present invention may further comprise antibodies or fragments thereof specific for other  
10       pathogens including bacteria, viral, parasitic, and fungal polypeptide sequences, in addition to the antibodies or fragments thereof of the present invention, for use in rapid differential pathogenic detection and diagnosis. The bio chips and biosensors of the present invention may also be used to monitor an *E. faecalis* or other Enterococcal infection and to monitor the genetic changes (amino acid deletions, insertions,  
15       substitutions, etc.) in response to drug therapy in the clinic and drug development in the laboratory. The bio chip and biosensors comprising polypeptides or antibodies of the present invention may also be used to simultaneously monitor the expression of a multiplicity of polypeptides, including those of the present invention. The polypeptides used to comprise a bio chip or biosensor of the present invention may be specified in the  
20       same manner as for the fragments, i.e., by their N-terminal and C-terminal positions or length in contiguous amino acid residue. Methods and particular uses of the polypeptides and antibodies of the present invention to detect Enterococcal species, including *E. faecalis*, or specific polypeptides using bio chip and biosensor technology include those known in the art, those of the U.S. Patent Nos. and World Patent Nos. listed above for  
25       bio chips and biosensors using polynucleotides of the present invention, and those of: U.S. Patent Nos. 5658732, 5135852, 5567301, 5677196, 5690894 and World Patent Nos. WO9729366, WO9612957, each incorporated herein in their entireties.

#### 4. Screening Assay for Binding Agents

30       Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Enterococcus faecalis* fragment and contigs herein described.

In general, such methods comprise steps of:

35       (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Enterococcus faecalis* genome; and

(b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides," in *Synthetic Peptides, A User's Guide*, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, *Biochemistry* 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991)) or to the mRNA itself (antisense - Okano, *J. Neurochem.* 56:560 (1991); *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences

of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

### 5. Pharmaceutical Compositions and Vaccines

5       The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present  
10       invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth and/or pathogenicity of *Enterococcus faecalis* or a related organism, *in vivo* or *in vitro*," when  
15       the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth by binding to an  
20       important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of a vaccine based on  
25       outer membrane components are well known in the art.

As used herein, a "related organism" is a broad term which refers to any organism whose growth can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related  
30       organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for  
35       treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most

cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical  
5 derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources,  
10 REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal  
15 stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.*, inhalation, intravenously,  
20 intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple  
25 injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from  
30 about 1  $\mu$ g/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each  
35 compound, or (2) the serum concentrations of each compound can be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a  
5 "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic  
10 administration of the compound(s) serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if  
15 its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known  
20 methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in a mixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, *e.g.*, human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16th Ed., Osol, A., Ed., Mack Publishing, Easton PA  
25 (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of  
30 action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or  
35 protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to

control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible  
5 to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques  
10 are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of  
15 pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

The present invention also provides vaccines comprising one or more  
20 polypeptides of the present invention. Heterogeneity in the composition of a vaccine may be provided by combining *E. faecalis* polypeptides of the present invention. Multi-component vaccines of this type are desirable because they are likely to be more effective in eliciting protective immune responses against multiple species and strains of the *Enterococcus* genus than single polypeptide vaccines.

25 Multi-component vaccines are known in the art to elicit antibody production to numerous immunogenic components. See, e.g., Decker et al. (1996) J. Infect. Dis. 174:S270-275. In addition, a hepatitis B, diphtheria, tetanus, pertussis tetravalent vaccine has recently been demonstrated to elicit protective levels of antibodies in human infants against all four pathogenic agents. See, e.g., Aristegui, J. et al. (1997) Vaccine  
30 15:7-9.

The present invention in addition to single-component vaccines includes multi-component vaccines. These vaccines comprise more than one polypeptide, immunogen or antigen. Thus, a multi-component vaccine would be a vaccine comprising more than one of the *E. faecalis* polypeptides of the present invention.

35 Further within the scope of the invention are whole cell and whole viral vaccines. Such vaccines may be produced recombinantly and involve the expression of one or more of the *E. faecalis* polypeptides described in SEQ ID NOS:1-982. For example, the *E.*

*faecalis* polypeptides of the present invention may be either secreted or localized intracellular, on the cell surface, or in the periplasmic space. Further, when a recombinant virus is used, the *E. faecalis* polypeptides of the present invention may, for example, be localized in the viral envelope, on the surface of the capsid, or internally within the capsid. Whole cells vaccines which employ cells expressing heterologous proteins are known in the art. See, e.g., Robinson, K. et al. (1997) Nature Biotech. 15:653-657; Sirard, J. et al. (1997) Infect. Immun. 65:2029-2033; Chabalgoity, J. et al. (1997) Infect. Immun. 65:2402-2412. These cells may be administered live or may be killed prior to administration. Chabalgoity, J. et al., *supra*, for example, report the successful use in mice of a live attenuated *Salmonella* vaccine strain which expresses a portion of a platyhelminth fatty acid-binding protein as a fusion protein on its cells surface.

A multi-component vaccine can also be prepared using techniques known in the art by combining one or more *E. faecalis* polypeptides of the present invention, or fragments thereof, with additional non-Enterococcal components (e.g., diphtheria toxin or tetanus toxin, and/or other compounds known to elicit an immune response). Such vaccines are useful for eliciting protective immune responses to both members of the *Enterococcus* genus and non-Enterococcal pathogenic agents.

The vaccines of the present invention also include DNA vaccines. DNA vaccines are currently being developed for a number of infectious diseases. See, et al., Boyer, et al. (1997) Nat. Med. 3:526-532; reviewed in Spier, R. (1996) Vaccine 14:1285-1288. Such DNA vaccines contain a nucleotide sequence encoding one or more *E. faecalis* polypeptides of the present invention oriented in a manner that allows for expression of the subject polypeptide. For example, the direct administration of plasmid DNA encoding *B. burgdorferi* OspA has been shown to elicit protective immunity in mice against borrelial challenge. See, Luke et al. (1997) J. Infect. Dis. 175:91-97.

The present invention also relates to the administration of a vaccine which is co-administered with a molecule capable of modulating immune responses. Kim et al. (1997) Nature Biotech. 15:641-646, for example, report the enhancement of immune responses produced by DNA immunizations when DNA sequences encoding molecules which stimulate the immune response are co-administered. In a similar fashion, the vaccines of the present invention may be co-administered with either nucleic acids encoding immune modulators or the immune modulators themselves. These immune modulators include granulocyte macrophage colony stimulating factor (GM-CSF) and CD86.

The vaccines of the present invention may be used to confer resistance to Enterococcal infection by either passive or active immunization. When the vaccines of

the present invention are used to confer resistance to Enterococcal infection through active immunization, a vaccine of the present invention is administered to an animal to elicit a protective immune response which either prevents or attenuates a Enterococcal infection. When the vaccines of the present invention are used to confer resistance to  
5 Enterococcal infection through passive immunization, the vaccine is provided to a host animal (e.g., human, dog, or mouse), and the antisera elicited by this antisera is recovered and directly provided to a recipient suspected of having an infection caused by a member of the *Enterococcus* genus.

The ability to label antibodies, or fragments of antibodies, with toxin molecules  
10 provides an additional method for treating Enterococcal infections when passive immunization is conducted. In this embodiment, antibodies, or fragments of antibodies, capable of recognizing the *E. faecalis* polypeptides disclosed herein, or fragments thereof, as well as other *Enterococcus* proteins, are labeled with toxin molecules prior to their administration to the patient. When such toxin derivatized antibodies bind to  
15 *Enterococcus* cells, toxin moieties will be localized to these cells and will cause their death.

The present invention thus concerns and provides a means for preventing or attenuating a Enterococcal infection resulting from organisms which have antigens that are recognized and bound by antisera produced in response to the polypeptides of the  
20 present invention. As used herein, a vaccine is said to prevent or attenuate a disease if its administration to an animal results either in the total or partial attenuation (i.e., suppression) of a symptom or condition of the disease, or in the total or partial immunity of the animal to the disease.

The administration of the vaccine (or the antisera which it elicits) may be for  
25 either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the compound(s) are provided in advance of any symptoms of Enterococcal infection. The prophylactic administration of the compound(s) serves to prevent or attenuate any subsequent infection. When provided therapeutically, the compound(s) is provided upon or after the detection of symptoms which indicate that an animal may be infected with a  
30 member of the *Enterococcus* genus. The therapeutic administration of the compound(s) serves to attenuate any actual infection. Thus, the *E. faecalis* polypeptides, and fragments thereof, of the present invention may be provided either prior to the onset of infection (so as to prevent or attenuate an anticipated infection) or after the initiation of an actual infection.

35 The polypeptides of the invention, whether encoding a portion of a native protein or a functional derivative thereof, may be administered in pure form or may be coupled to a macromolecular carrier. Example of such carriers are proteins and



carbohydrates. Suitable proteins which may act as macromolecular carrier for enhancing the immunogenicity of the polypeptides of the present invention include keyhole limpet hemacyanin (KLH) tetanus toxoid, pertussis toxin, bovine serum albumin, and ovalbumin. Methods for coupling the polypeptides of the present invention to such macromolecular carriers are disclosed in Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988).

A composition is said to be "pharmacologically or physiologically acceptable" if its administration can be tolerated by a recipient animal and is otherwise suitable for administration to that animal. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

While in all instances the vaccine of the present invention is administered as a pharmacologically acceptable compound, one skilled in the art would recognize that the composition of a pharmacologically acceptable compound varies with the animal to which it is administered. For example, a vaccine intended for human use will generally not be co-administered with Freund's adjuvant. Further, the level of purity of the *E. faecalis* polypeptides of the present invention will normally be higher when administered to a human than when administered to a non-human animal.

As would be understood by one of ordinary skill in the art, when the vaccine of the present invention is provided to an animal, it may be in a composition which may contain salts, buffers, adjuvants, or other substances which are desirable for improving the efficacy of the composition. Adjuvants are substances that can be used to specifically augment a specific immune response. These substances generally perform two functions: (1) they protect the antigen(s) from being rapidly catabolized after administration and (2) they nonspecifically stimulate immune responses.

Normally, the adjuvant and the composition are mixed prior to presentation to the immune system, or presented separately, but into the same site of the animal being immunized. Adjuvants can be loosely divided into several groups based upon their composition. These groups include oil adjuvants (for example, Freund's complete and incomplete), mineral salts (for example,  $AlK(SO_4)_2$ ,  $AlNa(SO_4)_2$ ,  $AlNH_4(SO_4)$ , silica, kaolin, and carbon), polynucleotides (for example, poly IC and poly AU acids), and certain natural substances (for example, wax D from *Mycobacterium tuberculosis*, as well as substances found in *Corynebacterium parvum*, or *Bordetella pertussis*, and members of the genus *Brucella*). Other substances useful as adjuvants are the saponins such as, for example, Quil A. (Superfos A/S, Denmark). Preferred adjuvants for use in the present invention include aluminum salts, such as  $AlK(SO_4)_2$ ,  $AlNa(SO_4)_2$ , and  $AlNH_4(SO_4)$ .

Examples of materials suitable for use in vaccine compositions are provided in REMINGTON'S PHARMACEUTICAL SCIENCES 1324-1341 (A. Osol, ed, Mack Publishing Co, Easton, PA, (1980) (incorporated herein by reference).

The therapeutic compositions of the present invention can be administered  
5 parenterally by injection, rapid infusion, nasopharyngeal absorption  
(intranasopharyngeally), dermoabsorption, or orally. The compositions may  
alternatively be administered intramuscularly, or intravenously. Compositions for  
parenteral administration include sterile aqueous or non-aqueous solutions, suspensions,  
and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene  
10 glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate.  
Carriers or occlusive dressings can be used to increase skin permeability and enhance  
antigen absorption. Liquid dosage forms for oral administration may generally comprise a  
liposome solution containing the liquid dosage form. Suitable forms for suspending  
liposomes include emulsions, suspensions, solutions, syrups, and elixirs containing inert  
15 diluents commonly used in the art, such as purified water. Besides the inert diluents, such  
compositions can also include adjuvants, wetting agents, emulsifying and suspending  
agents, or sweetening, flavoring, or perfuming agents.

Therapeutic compositions of the present invention can also be administered in  
encapsulated form. For example, intranasal immunization using vaccines encapsulated in  
20 biodegradable microsphere composed of poly(DL-lactide-co-glycolide). See, Shahin, R. et  
al. (1995) Infect. Immun. 63:1195-1200. Similarly, orally administered encapsulated  
*Salmonella typhimurium* antigens can also be used. Allaoui-Attarki, K. et al. (1997)  
Infect. Immun. 65:853-857. Encapsulated vaccines of the present invention can be  
administered by a variety of routes including those involving contacting the vaccine with  
25 mucous membranes (e.g., intranasally, intracolonicly, intraduodenally).

Many different techniques exist for the timing of the immunizations when a  
multiple administration regimen is utilized. It is possible to use the compositions of the  
invention more than once to increase the levels and diversities of expression of the  
immunoglobulin repertoire expressed by the immunized animal. Typically, if multiple  
30 immunizations are given, they will be given one to two months apart.

According to the present invention, an "effective amount" of a therapeutic  
composition is one which is sufficient to achieve a desired biological effect. Generally,  
the dosage needed to provide an effective amount of the composition will vary depending  
upon such factors as the animal's or human's age, condition, sex, and extent of disease, if  
35 any, and other variables which can be adjusted by one of ordinary skill in the art.

The antigenic preparations of the invention can be administered by either single  
or multiple dosages of an effective amount. Effective amounts of the compositions of

the invention can vary from 0.01-1,000  $\mu\text{g/ml}$  per dose, more preferably 0.1-500  $\mu\text{g/ml}$  per dose, and most preferably 10-300  $\mu\text{g/ml}$  per dose.

#### 6. Shot-Gun Approach to Megabase DNA Sequencing

5 The present invention further demonstrates that a large genome can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

10 Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

### ILLUSTRATIVE EXAMPLES

15

#### LIBRARIES AND SEQUENCING

##### 1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics* 2:231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability,  $P_0$ , that any given base in a sequence of size  $L$ , in nucleotides, is not sequenced after a certain amount,  $n$ , in nucleotides, of random sequence has been determined can be calculated by the equation  $P_0 = e^{-m}$ , where  $m$  is  $L/n$ , the fold coverage. For instance, for a genome of 2.8 Mb,  $m=1$  when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point,  $P_0 = e^{-1} = 0.37$ . The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence  $L$  has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size  $L$ , in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

35 Similarly, the total gap length,  $G$ , is determined by the equation  $G = Le^{-m}$ , and the average gap size,  $g$ , follows the equation,  $g = L/n$ . Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

## 2. Random Library Construction

5 In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

*Enterococcus faecalis* DNA is prepared by phenol extraction. A mixture containing 200 µg DNA in 1.0 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 50% glycerol is processed through a nebulizer (IPI Medical Products) with a stream of nitrogen adjusted to 35 Kpa for 2 minutes. The sonicated DNA is ethanol precipitated and redissolved in 500 µl TE buffer.

To create blunt-ends, a 100 µl aliquot of the resuspended DNA is digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 µl BAL31 buffer. The digested DNA is phenol-extracted, ethanol-precipitated, redissolved in 100 µl TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size is excised from the gel, and the LGT agarose is melted and the resulting solution is extracted with phenol to separate the agarose from the DNA. DNA is ethanol precipitated and redissolved in 20 µl of TE buffer for ligation to vector.

A two-step ligation procedure is used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 µl) contains 2 µg of DNA fragments, 2 µg pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and is incubated at 14°C for 4 hr. The ligation mixture then is phenol extracted and ethanol precipitated, and the precipitated DNA is dissolved in 20 µl TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder are visualized by ethidium bromide-staining and UV illumination and identified by size as insert (I), vector (v), v+I, v+2i, v+3i, etc. The portion of the gel containing v+I DNA is excised and the v+I DNA is recovered and resuspended into 20 µl TE. The v+I DNA then is blunt-ended by T4 polymerase treatment for 5 min. at 37°C in a reaction mixture (50 µl) containing the v+I linears, 500 µM each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+I linears are dissolved in 20 µl TE.

35 The final ligation to produce circles is carried out in a 50 µl reaction containing 5 µl of v+I linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture is stored at -20°C.

This two-stage procedure results in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E. coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies* 3 (1):5 (1990)) are used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells are plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating is carried out as follows. A 100 µl aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) is thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 µl aliquot of 1.42 M beta-mercaptoethanol is added to the aliquot of cells to a final concentration of 25 mM. Cells are incubated on ice for 10 min. A 1 µl aliquot of the final ligation is added to the cells and incubated on ice for 30 min. The cells are heat pulsed for 30 sec. at 42°C and placed back on ice for 2 min. The outgrowth period in liquid culture is eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture is plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl<sub>2</sub> (1 M), and 1 ml MgSO<sub>4</sub>/100 ml SOB agar. The 15 ml top layer is poured just prior to plating. Our titer is approximately 100 colonies/10 µl aliquot of transformation.

All colonies are picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products are deleted from the library, resulting in a slight increase in gap number over that expected.

### 3. Random DNA Sequencing

High quality double stranded DNA plasmid templates are prepared using a "boiling bead" method developed in collaboration with Advanced Genetic Technology Corp. (Gaithersburg, MD) (Adams *et al.*, *Science* 252:1651 (1991); Adams *et al.*, *Nature* 355:632 (1992)). Plasmid preparation is performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Template concentration is determined using Hoechst Dye and a Millipore Cytofluor. DNA concentrations are not adjusted, but low-yielding templates are identified where possible and not sequenced.

Templates are also prepared from an *Enterococcus faecalis* lambda genomic library in the vector DASH II (Stratagene). In particular, *Enterococcus faecalis* DNA (> 100 kb) is partially digested in a reaction mixture (200  $\mu$ l) containing 50  $\mu$ g DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23°C. The digested DNA was phenol-  
5 extracted and fractionated by sucrose density gradient centrifugation. Fractions of the sucrose gradient containing 15 to 25 kb are recovered in a final volume of 6  $\mu$ l. One  $\mu$ l of fragments is used with 1  $\mu$ l of lambda DASHII vector (Stratagene) in the recommended ligation reaction. One  $\mu$ l of the ligation mixture is used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract (Stratagene,  
10 #227711). Phage are plated directly without amplification from the packaging mixture (after dilution with 500  $\mu$ l of recommended SM buffer and chloroform treatment). Yield is about  $2.5 \times 10^3$  pfu/ $\mu$ l. An amplified library is prepared by infecting restructure NM539 host *E. coli* cells with approximately  $1 \times 10^4$  phage particles and recovering the progeny phages particles. The recovered phage is stored frozen in 7% dimethylsulfoxide. The  
15 phage titer is approximately  $1 \times 10^9$  pfu/ml.

For high throughput sequencing of individual lambda phage clones, liquid lysates (100  $\mu$ l) are prepared from randomly selected plaques (from the unamplified library) and template is prepared by long-range PCR using T7 and T3 vector-specific primers.

Sequencing reactions are carried out on plasmid and/or PCR templates using the  
20 AB Catalyst LabStation with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers (Adams *et al.*, *Nature* 368:474 (1994)). Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. T7 and T3  
25 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are performed by eight individuals using an average of fourteen AB 373 DNA Sequencers per day. All sequencing reactions are analyzed using the Stretch modification of the AB 373, primarily using a 34 cm well-to-read distance. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1  
30 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

Richards *et al.*, Chapter 28 in AUTOMATED DNA SEQUENCING AND ANALYSIS, M. D. Adams, C. Fields, J. C. Venter, Eds., Academic Press, London, (1994)  
35 described the value of using sequence from both ends of sequencing templates to facilitate ordering of contigs in shotgun assembly projects of lambda and cosmid clones. We balance the desirability of both-end sequencing (including the reduced cost of lower total

number of templates) against shorter read-lengths for sequencing reactions performed with the M13RP1 (reverse) primer compared to the M13-21 (forward) primer.

Approximately one-half of the templates are sequenced from both ends. Random reverse sequencing reactions are done based on successful forward sequencing reactions. Some  
5 M13RP1 sequences are obtained in a semi-directed fashion: M13-21: sequences pointing outward at the ends of contigs are chosen for M13RP1 sequencing in an effort to specifically order contigs.

#### 4. Protocol for Automated Cycle Sequencing

10 The sequencing was carried out using ABI Catalyst robots and AB 373 Automated DNA Sequencers. The Catalyst robot is a publicly available sophisticated pipetting and temperature control robot which has been developed specifically for DNA sequencing reactions. The Catalyst combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-  
15 labelled sequencing primers, and reaction buffer. Reaction mixes and templates are combined in the wells of an aluminum 96-well thermocycling plate. Thirty consecutive cycles of linear amplification (*i.e.*, one primer synthesis) steps are performed including denaturation, annealing of primer and template, and extension; *i.e.*, DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without  
20 the need for an oil overlay.

Two sequencing protocols are used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer is labelled with a different fluorescent dye, permitting the four individual reactions  
25 to be combined into one lane of the 373 DNA Sequencer for electrophoresis, detection, and base-calling. ABI currently supplies pre-mixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR- generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give  
30 longer usable sequences.

Thirty-two reactions are loaded per AB373 Sequencer each day, for a total of 960 samples. Electrophoresis is run overnight following the manufacturer's protocols, and the data is collected for twelve hours. Following electrophoresis and fluorescence detection, the ABI 373 performs automatic lane tracking and base-calling. The lane-tracking is  
35 confirmed visually. Each sequence electropherogram (or fluorescence lane trace) is inspected visually and assessed for quality. Trailing sequences of low quality are removed and the sequence itself is loaded via software to a Sybase database (archived daily to 8mm

tape). Leading vector polylinker sequence is removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 are around 400 bp and depend mostly on the quality of the template used for the sequencing reaction. ABI 373 Sequencers converted to Stretch Liners provide a longer electrophoresis path  
5 prior to fluorescence detection and increase the average number of usable bases to 500-600 bp.

## INFORMATICS

### 1. Data Management

10 A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database  
15 management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen is based on a Unix platform, it was necessary to design and  
20 implement a variety of multi- user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

### 2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate  
25 assembly of thousands of sequence fragments is employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than 104 fragments, the algorithm builds a hash table of 10 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines  
30 which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology*  
35 164:765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These



criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain range of base pairs (definable for each clone based on the known clone size range for a given library).

The process resulted in 982 contigs as represented by SEQ ID NOs:1-982.

### 3. Identifying Genes

The predicted coding regions of the *Enterococcus faecalis* genome were initially defined with the program GeneMark, which finds ORFs using a probabilistic classification technique. The predicted coding region sequences were used in searches against a database of all *Enterococcus faecali* nucleotide sequences from GenBank (March, 1997), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases. ORFs that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

## ILLUSTRATIVE APPLICATIONS

### 1. Production of an Antibody to a *Enterococcus faecalis* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

## 2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.*, *Basic Methods in Molecular Biology*, Elsevier, New York. Section 21-2 (1989).

## 3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as

described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980)

Antibody preparations prepared according to either protocol are useful in  
5 quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi- quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, antibodies are useful in various animal models of enterococcal disease as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a  
10 potential immunotherapeutic or immunoprophylactic reagent.

#### 4. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Enterococcus faecalis* genome, such as those of Tables 1-3 and SEQ ID NOS:1-982 can be used, in accordance with the present invention, to  
15 prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

20

#### 5. Isolation of a Selected DNA Clone From the Deposited

##### Sample of *E. faecalis*

Three approaches can be used to isolate a *E. faecalis* clone comprising a polynucleotide of the present invention from any *E. faecalis* genomic DNA library. The  
25 *E. faecalis* strain V586 has been deposited as a convenient source for obtaining a *E. faecalis* strain although a wide variety of strains *E. faecalis* strains can be used which are known in the art.

*E. faecalis* genomic DNA is prepared using the following method. A 20ml overnight bacterial culture grown in a rich medium (e.g., Trypticase Soy Broth, Brain  
30 Heart Infusion broth or Super broth), pelleted, washed two times with TES (30mM Tris-pH 8.0, 25mM EDTA, 50mM NaCl), and resuspended in 5ml high salt TES (2.5M NaCl). Lysostaphin is added to final concentration of approx 50ug/ml and the mixture is rotated slowly 1 hour at 37C to make protoplast cells. The solution is then placed in incubator (or place in a shaking water bath) and warmed to 55C. Five hundred micro liter of 20%  
35 sarcosyl in TES (final concentration 2%) is then added to lyse the cells. Next, guanidine HCl is added to a final concentration of 7M (3.69g in 5.5 ml). The mixture is swirled slowly at 55C for 60-90 min (solution should clear). A CsCl gradient is then set up in

SW41 ultra clear tubes using 2.0ml 5.7M CsCl and overlaying with 2.85M CsCl. The gradient is carefully overlayed with the DNA-containing GuHCl solution. The gradient is spun at 30,000 rpm, 20C for 24 hr and the lower DNA band is collected. The volume is increased to 5 ml with TE buffer. The DNA is then treated with protease K (10 ug/ml) overnight at 37 C, and precipitated with ethanol. The precipitated DNA is resuspended in a desired buffer.

In the first method, a plasmid is directly isolated by screening a plasmid *E. faecalis* genomic DNA library using a polynucleotide probe corresponding to a polynucleotide of the present invention. Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with  $^{32}\text{P}$ - $\gamma$ -ATP using T4 polynucleotide kinase and purified according to routine methods. (See, e.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The library is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art. See, e.g., Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL (Cold Spring Harbor, N.Y. 2nd ed. 1989); Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (John Wiley and Sons, N.Y. 1989). The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening. See, e.g., Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL (Cold Spring Harbor, N.Y. 2nd ed. 1989); Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (John Wiley and Sons, N.Y. 1989) or other techniques known to those of skill in the art.

Alternatively, two primers of 15-25 nucleotides derived from the 5' and 3' ends of a polynucleotide of SEQ ID NOS:1-982 are synthesized and used to amplify the desired DNA by PCR using a *E. faecalis* genomic DNA prep as a template. PCR is carried out under routine conditions, for instance, in 25  $\mu\text{l}$  of reaction mixture with 0.5  $\mu\text{g}$  of the above DNA template. A convenient reaction mixture is 1.5-5 mM  $\text{MgCl}_2$ , 0.01% (w/v) gelatin, 20  $\mu\text{M}$  each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the

DNA product.

Finally, overlapping oligos of the DNA sequences of SEQ ID NOS:1-982 can be chemically synthesized and used to generate a nucleotide sequence of desired length using PCR methods known in the art.

5

**6(a). Expression and Purification Enterococcal polypeptides  
in *E. coli***

The bacterial expression vector pQE60 was used for bacterial expression of some of the polypeptide fragments of the present invention which were used in the soft tissue and systemic infection models discussed below. (QIAGEN, Inc., 9259 Eton Avenue,  
10 Chatsworth, CA, 91311). pQE60 encodes ampicillin antibiotic resistance ("Ampr") and contains a bacterial origin of replication ("ori"), an IPTG inducible promoter, a ribosome binding site ("RBS"), six codons encoding histidine residues that allow affinity purification using nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin (QIAGEN, Inc., *supra*) and  
15 suitable single restriction enzyme cleavage sites. These elements are arranged such that an inserted DNA fragment encoding a polypeptide expresses that polypeptide with the six His residues (i.e., a "6 X His tag") covalently linked to the carboxyl terminus of that polypeptide.

The DNA sequence encoding the desired portion of a *E. faecalis* protein of the present invention was amplified from *E. faecalis* genomic DNA using PCR  
20 oligonucleotide primers which anneal to the 5' and 3' sequences coding for the portions of the *E. faecalis* polynucleotide shown in SEQ ID NOS:1-982. Additional nucleotides containing restriction sites to facilitate cloning in the pQE60 vector are added to the 5' and 3' sequences, respectively.

For cloning the mature protein, the 5' primer has a sequence containing an appropriate restriction site followed by nucleotides of the amino terminal coding sequence of the desired *E. faecalis* polynucleotide sequence in SEQ ID NOS:1-982. One of ordinary skill in the art would appreciate that the point in the protein coding sequence where the 5' and 3' primers begin may be varied to amplify a DNA segment encoding any desired  
30 portion of the complete protein shorter or longer than the mature form. The 3' primer has a sequence containing an appropriate restriction site followed by nucleotides complementary to the 3' end of the polypeptide coding sequence of SEQ ID NOS:1-982, excluding a stop codon, with the coding sequence aligned with the restriction site so as to maintain its reading frame with that of the six His codons in the pQE60 vector.

35 The amplified *E. faecalis* DNA fragment and the vector pQE60 were digested with restriction enzymes which recognize the sites in the primers and the digested DNAs were then ligated together. The *E. faecalis* DNA was inserted into the restricted pQE60

vector in a manner which places the *E. faecalis* protein coding region downstream from the IPTG-inducible promoter and in-frame with an initiating AUG and the six histidine codons.

The ligation mixture was transformed into competent *E. coli* cells using standard procedures such as those described by Sambrook et al., *supra*. *E. coli* strain M15/rep4, containing multiple copies of the plasmid pREP4, which expresses the lac repressor and confers kanamycin resistance ("Kanr"), was used in carrying out the illustrative example described herein. This strain, which was only one of many that are suitable for expressing a *E. faecalis* polypeptide, is available commercially (QIAGEN, Inc., *supra*). Transformants were identified by their ability to grow on LB agar plates in the presence of ampicillin and kanamycin. Plasmid DNA was isolated from resistant colonies and the identity of the cloned DNA confirmed by restriction analysis, PCR and DNA sequencing.

Clones containing the desired constructs were grown overnight ("O/N") in liquid culture in LB media supplemented with both ampicillin (100 µg/ml) and kanamycin (25 µg/ml). The O/N culture was used to inoculate a large culture, at a dilution of approximately 1:25 to 1:250. The cells were grown to an optical density at 600 nm ("OD600") of between 0.4 and 0.6. Isopropyl-β-D-thiogalactopyranoside ("IPTG") was then added to a final concentration of 1 mM to induce transcription from the lac repressor sensitive promoter, by inactivating the lacI repressor. Cells subsequently were incubated further for 3 to 4 hours. Cells then were harvested by centrifugation.

The cells were then stirred for 3-4 hours at 4°C in 6M guanidine-HCl, pH 8. The cell debris was removed by centrifugation, and the supernatant containing the *E. faecalis* polypeptide was loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity were purified in a simple one-step procedure (for details see: The QIAexpressionist, 1995, QIAGEN, Inc., *supra*). Briefly the supernatant was loaded onto the column in 6 M guanidine-HCl, pH 8, the column was first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the *E. faecalis* polypeptide was eluted with 6 M guanidine-HCl, pH 5.

The purified protein was then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein could be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins can be eluted by the addition of 250 mM imidazole. Imidazole was removed by a final dialyzing step against PBS or 50 mM sodium acetate

pH 6 buffer plus 200 mM NaCl. The purified protein was stored at 4° C or frozen at -80° C.

Some of the polypeptide of the present invention were prepared using a non-denaturing protein purification method. For these polypeptides, the cell pellet from each liter of culture was resuspended in 25 mls of Lysis Buffer A at 4°C (Lysis Buffer A = 50 mM Na-phosphate, 300 mM NaCl, 10 mM 2-mercaptoethanol, 10% Glycerol, pH 7.5 with 1 tablet of Complete EDTA-free protease inhibitor cocktail (Boehringer Mannheim #1873580) per 50 ml of buffer). Absorbance at 550 nm was approximately 10-20 O.D./ml. The suspension was then put through three freeze/thaw cycles from -70°C (using a ethanol-dry ice bath) up to room temperature. The cells were lysed via sonication in short 10 sec bursts over 3 minutes at approximately 80W while kept on ice. The sonicated sample was then centrifuged at 15,000 RPM for 30 minutes at 4°C. The supernatant was passed through a column containing 1.0 ml of CL-4B resin to pre-clear the sample of any proteins that may bind to agarose non-specifically, and the flow-through fraction was collected.

The pre-cleared flow-through was applied to a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (Quiagen, Inc., *supra*). Proteins with a 6 X His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure. Briefly, the supernatant was loaded onto the column in Lysis Buffer A at 4°C, the column was first washed with 10 volumes of Lysis Buffer A until the A280 of the eluate returns to the baseline. Then, the column was washed with 5 volumes of 40 mM Imidazole (92% Lysis Buffer A / 8% Buffer B) (Buffer B = 50 mM Na-Phosphate, 300 mM NaCl, 10% Glycerol, 10 mM 2-mercaptoethanol, 500 mM Imidazole, pH of the final buffer should be 7.5). The protein was eluted off of the column with a series of increasing Imidazole solutions made by adjusting the ratios of Lysis Buffer A to Buffer B. Three different concentrations were used: 3 volumes of 75 mM Imidazole, 3 volumes of 150 mM Imidazole, 5 volumes of 500 mM Imidazole. The fractions containing the purified protein were analyzed using 8 %, 10 % or 14% SDS-PAGE depending on the protein size. The purified protein was then dialyzed 2X against phosphate-buffered saline (PBS) in order to place it into an easily workable buffer. The purified protein was stored at 4° C or frozen at -80°.

The following alternative method may be used to purify *E. faecalis* expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells are harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit

weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

5       The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 x g for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

10       The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 x g centrifugation for 15 min., the pellet is discarded and the *E. faecalis* polypeptide-containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 x g) to remove insoluble particles,  
15       the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded *E. faecalis* polypeptide solution, a previously prepared  
20       tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise  
25       manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the *E. faecalis* polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion  
30       (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A<sub>280</sub>  
35       monitoring of the effluent. Fractions containing the *E. faecalis* polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.



The resultant *E. faecalis* polypeptide exhibits greater than 95% purity after the above refolding and purification steps. No major contaminant bands are observed from Commassie blue stained 16% SDS-PAGE gel when 5 µg of purified protein is loaded. The purified protein is also tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

#### **6(b). Alternative Expression and Purification Enterococcal polypeptides in *E. coli***

The vector pQE10 was alternatively used to clone and express some of the polypeptides of the present invention for use in the soft tissue and systemic infection models discussed below. The difference being such that an inserted DNA fragment encoding a polypeptide expresses that polypeptide with the six His residues (i.e., a "6 X His tag") covalently linked to the amino terminus of that polypeptide. The bacterial expression vector pQE10 (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311) was used in this example. The components of the pQE10 plasmid are arranged such that the inserted DNA sequence encoding a polypeptide of the present invention expresses the polypeptide with the six His residues (i.e., a "6 X His tag")) covalently linked to the amino terminus.

The DNA sequences encoding the desired portions of a polypeptide of SEQ ID NOS:1-982 were amplified using PCR oligonucleotide primers from genomic *E. faecalis* DNA. The PCR primers anneal to the nucleotide sequences encoding the desired amino acid sequence of a polypeptide of the present invention. Additional nucleotides containing restriction sites to facilitate cloning in the pQE10 vector were added to the 5' and 3' primer sequences, respectively.

For cloning a polypeptide of the present invention, the 5' and 3' primers were selected to amplify their respective nucleotide coding sequences. One of ordinary skill in the art would appreciate that the point in the protein coding sequence where the 5' and 3' primers begins may be varied to amplify a DNA segment encoding any desired portion of a polypeptide of the present invention. The 5' primer was designed so the coding sequence of the 6 X His tag is aligned with the restriction site so as to maintain its reading frame with that of *E. faecalis* polypeptide. The 3' was designed to include an stop codon. The amplified DNA fragment was then cloned, and the protein expressed, as described above for the pQE60 plasmid.

The DNA sequences encoding the amino acid sequences of SEQ ID NOS:1-982 may also be cloned and expressed as fusion proteins by a protocol similar to that described directly above, wherein the pET-32b(+) vector (Novagen, 601 Science Drive, Madison, WI 53711) is preferentially used in place of pQE10.

The above methods are not limited to the polypeptide fragments actually produced. The above method, like the methods below, can be used to produce either full length polypeptides or desired fragments thereof.

5           **6(c). Alternative Expression and Purification of Enterococcal polypeptides in *E. coli***

The bacterial expression vector pQE60 is used for bacterial expression in this example (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311). However, in this example, the polypeptide coding sequence is inserted such that translation of the six His  
10       codons is prevented and, therefore, the polypeptide is produced with no 6 X His tag.

The DNA sequence encoding the desired portion of the *E. faecalis* amino acid sequence is amplified from an *E. faecalis* genomic DNA prep the deposited DNA clones using PCR oligonucleotide primers which anneal to the 5' and 3' nucleotide sequences corresponding to the desired portion of the *E. faecalis* polypeptides. Additional  
15       nucleotides containing restriction sites to facilitate cloning in the pQE60 vector are added to the 5' and 3' primer sequences.

For cloning a *E. faecalis* polypeptides of the present invention, 5' and 3' primers are selected to amplify their respective nucleotide coding sequences. One of ordinary skill in the art would appreciate that the point in the protein coding sequence where the 5' and  
20       3' primers begin may be varied to amplify a DNA segment encoding any desired portion of a polypeptide of the present invention. The 3' and 5' primers contain appropriate restriction sites followed by nucleotides complementary to the 5' and 3' ends of the coding sequence respectively. The 3' primer is additionally designed to include an in-frame stop codon.

25       The amplified *E. faecalis* DNA fragments and the vector pQE60 are digested with restriction enzymes recognizing the sites in the primers and the digested DNAs are then ligated together. Insertion of the *E. faecalis* DNA into the restricted pQE60 vector places the *E. faecalis* protein coding region including its associated stop codon downstream from the IPTG-inducible promoter and in-frame with an initiating AUG.  
30       The associated stop codon prevents translation of the six histidine codons downstream of the insertion point.

The ligation mixture is transformed into competent *E. coli* cells using standard procedures such as those described by Sambrook et al. *E. coli* strain M15/rep4, containing multiple copies of the plasmid pREP4, which expresses the lac repressor and confers  
35       kanamycin resistance ("Kanr"), is used in carrying out the illustrative example described herein. This strain, which is only one of many that are suitable for expressing *E. faecalis* polypeptide, is available commercially (QIAGEN, Inc., *supra*). Transformants are

identified by their ability to grow on LB plates in the presence of ampicillin and kanamycin. Plasmid DNA is isolated from resistant colonies and the identity of the cloned DNA confirmed by restriction analysis, PCR and DNA sequencing.

Clones containing the desired constructs are grown overnight ("O/N") in liquid  
5 culture in LB media supplemented with both ampicillin (100 µg/ml) and kanamycin (25 µg/ml). The O/N culture is used to inoculate a large culture, at a dilution of approximately 1:25 to 1:250. The cells are grown to an optical density at 600 nm ("OD600") of between 0.4 and 0.6. isopropyl-b-D-thiogalactopyranoside ("IPTG") is then added to a final concentration of 1 mM to induce transcription from the *lac* repressor sensitive  
10 promoter, by inactivating the *lacI* repressor. Cells subsequently are incubated further for 3 to 4 hours. Cells then are harvested by centrifugation.

To purify the *E. faecalis* polypeptide, the cells are then stirred for 3-4 hours at 4°C in 6M guanidine-HCl, pH 8. The cell debris is removed by centrifugation, and the supernatant containing the *E. faecalis* polypeptide is dialyzed against 50 mM Na-acetate  
15 buffer pH 6, supplemented with 200 mM NaCl. Alternatively, the protein can be successfully refolded by dialyzing it against 500 mM NaCl, 20% glycerol, 25 mM Tris/HCl pH 7.4, containing protease inhibitors. After renaturation the protein can be purified by ion exchange, hydrophobic interaction and size exclusion chromatography. Alternatively, an affinity chromatography step such as an antibody column can be used to  
20 obtain pure *E. faecalis* polypeptide. The purified protein is stored at 4°C or frozen at -80°C.

The following alternative method may be used to purify *E. faecalis* polypeptides expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

25 Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells are harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM  
30 EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells were then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by  
35 centrifugation at 7000 x g for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 x g centrifugation for 15 min., the pellet is discarded and the *E. faecalis* polypeptide-containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

5        Following high speed centrifugation (30,000 x g) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

10        To clarify the refolded *E. faecalis* polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with  
15        250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the *E. faecalis* polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of  
20        tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0  
25        M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A<sub>280</sub> monitoring of the effluent. Fractions containing the *E. faecalis* polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant *E. faecalis* polypeptide exhibits greater than 95% purity after the above refolding and purification steps. No major contaminant bands are observed from  
30        Commassie blue stained 16% SDS-PAGE gel when 5 µg of purified protein is loaded. The purified protein is also tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

#### 6(d). Cloning and Expression of *E. faecalis* in Other Bacteria

35        *E. faecalis* polypeptides can also be produced in: *E. faecalis* using the methods of S. Skinner et al., (1988) Mol. Microbiol. 2:289-297 or J. I. Moreno (1996) Protein Expr.

Purif. 8(3):332-340; *Lactobacillus* using the methods of C. Rush et al., 1997 Appl. Microbiol. Biotechnol. 47(5):537-542; or in *Bacillus subtilis* using the methods Chang et al., U.S. Patent No. 4,952,508.

## 5           7. Cloning and Expression in COS Cells

A *E. faecalis* expression plasmid is made by cloning a portion of the DNA encoding a *E. faecalis* polypeptide into the expression vector pDNAI/Amp or pDNAIII (which can be obtained from Invitrogen, Inc.). The expression vector pDNAI/amp contains: (1) an *E. coli* origin of replication effective for propagation in *E. coli* and other  
10 prokaryotic cells; (2) an ampicillin resistance gene for selection of plasmid-containing prokaryotic cells; (3) an SV40 origin of replication for propagation in eukaryotic cells; (4) a CMV promoter, a polylinker, an SV40 intron; (5) several codons encoding a hemagglutinin fragment (i.e., an "HA" tag to facilitate purification) followed by a termination codon and polyadenylation signal arranged so that a DNA can be  
15 conveniently placed under expression control of the CMV promoter and operably linked to the SV40 intron and the polyadenylation signal by means of restriction sites in the polylinker. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein described by Wilson et al. 1984 Cell 37:767. The fusion of the HA tag to the target protein allows easy detection and recovery of the recombinant protein  
20 with an antibody that recognizes the HA epitope. pDNAIII contains, in addition, the selectable neomycin marker.

A DNA fragment encoding a *E. faecalis* polypeptide is cloned into the polylinker region of the vector so that recombinant protein expression is directed by the CMV promoter. The plasmid construction strategy is as follows. The DNA from a *E. faecalis*  
25 genomic DNA prep is amplified using primers that contain convenient restriction sites, much as described above for construction of vectors for expression of *E. faecalis* in *E. coli*. The 5' primer contains a Kozak sequence, an AUG start codon, and nucleotides of the 5' coding region of the *E. faecalis* polypeptide. The 3' primer, contains nucleotides complementary to the 3' coding sequence of the *E. faecalis* DNA, a stop  
30 codon, and a convenient restriction site.

The PCR amplified DNA fragment and the vector, pDNAI/Amp, are digested with appropriate restriction enzymes and then ligated. The ligation mixture is transformed into an appropriate *E. coli* strain such as SURE™ (Stratagene Cloning Systems, La Jolla, CA 92037), and the transformed culture is plated on ampicillin media plates which then  
35 are incubated to allow growth of ampicillin resistant colonies. Plasmid DNA is isolated from resistant colonies and examined by restriction analysis or other means for the presence of the fragment encoding the *E. faecalis* polypeptide

For expression of a recombinant *E. faecalis* polypeptide, COS cells are transfected with an expression vector, as described above, using DEAE-dextran, as described, for instance, by Sambrook et al. (*supra*). Cells are incubated under conditions for expression of *E. faecalis* by the vector.

5        Expression of the *E. faecalis*-HA fusion protein is detected by radiolabeling and immunoprecipitation, using methods described in, for example Harlow et al., *supra*.. To this end, two days after transfection, the cells are labeled by incubation in media containing <sup>35</sup>S-cysteine for 8 hours. The cells and the media are collected, and the cells are washed and the lysed with detergent-containing RIPA buffer: 150 mM NaCl, 1% NP-  
10    40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50 mM TRIS, pH 7.5, as described by Wilson et al. (*supra* ). Proteins are precipitated from the cell lysate and from the culture media using an HA-specific monoclonal antibody. The precipitated proteins then are analyzed by SDS-PAGE and autoradiography. An expression product of the expected size is seen in the cell lysate, which is not seen in negative controls.

15

#### 8. Cloning and Expression in CHO Cells

The vector pC4 is used for the expression of *E. faecalis* polypeptide in this example. Plasmid pC4 is a derivative of the plasmid pSV2-dhfr (ATCC Accession No. 37146). The plasmid contains the mouse DHFR gene under control of the SV40 early  
20    promoter. Chinese hamster ovary cells or other cells lacking dihydrofolate activity that are transfected with these plasmids can be selected by growing the cells in a selective medium (alpha minus MEM, Life Technologies) supplemented with the chemotherapeutic agent methotrexate. The amplification of the DHFR genes in cells resistant to methotrexate (MTX) has been well documented. See, e.g., Alt et al., 1978, J. Biol.  
25    Chem. 253:1357-1370; Hamlin et al., 1990, Biochem. et Biophys. Acta, 1097:107-143; Page et al., 1991, Biotechnology 9:64-68. Cells grown in increasing concentrations of MTX develop resistance to the drug by overproducing the target enzyme, DHFR, as a result of amplification of the DHFR gene. If a second gene is linked to the DHFR gene, it is usually co-amplified and over-expressed. It is known in the art that this approach may  
30    be used to develop cell lines carrying more than 1,000 copies of the amplified gene(s). Subsequently, when the methotrexate is withdrawn, cell lines are obtained which contain the amplified gene integrated into one or more chromosome(s) of the host cell.

Plasmid pC4 contains the strong promoter of the long terminal repeat (LTR) of the Rouse Sarcoma Virus, for expressing a polypeptide of interest, Cullen, et al. (1985)  
35    Mol. Cell. Biol. 5:438-447; plus a fragment isolated from the enhancer of the immediate early gene of human cytomegalovirus (CMV), Boshart, et al., 1985, Cell 41:521-530. Downstream of the promoter are the following single restriction enzyme cleavage sites

that allow the integration of the genes: *Bam* HI, *Xba* I, and *Asp* 718. Behind these cloning sites the plasmid contains the 3' intron and polyadenylation site of the rat preproinsulin gene. Other high efficiency promoters can also be used for the expression, e.g., the human  $\beta$ -actin promoter, the SV40 early or late promoters or the long terminal repeats from other retroviruses, e.g., HIV and HTLV1. Clontech's Tet-Off and Tet-On gene expression systems and similar systems can be used to express the *E. faecalis* polypeptide in a regulated way in mammalian cells (Gossen et al., 1992, Proc. Natl. Acad. Sci. USA 89:5547-5551. For the polyadenylation of the mRNA other signals, e.g., from the human growth hormone or globin genes can be used as well. Stable cell lines carrying a gene of interest integrated into the chromosomes can also be selected upon co-  
transfection with a selectable marker such as gpt, G418 or hygromycin. It is advantageous to use more than one selectable marker in the beginning, e.g., G418 plus methotrexate.

The plasmid pC4 is digested with the restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel. The DNA sequence encoding the *E. faecalis* polypeptide is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the desired portion of the gene. A 5' primer containing a restriction site, a Kozak sequence, an AUG start codon, and nucleotides of the 5' coding region of the *E. faecalis* polypeptide is synthesized and used. A 3' primer, containing a restriction site, stop codon, and nucleotides complementary to the 3' coding sequence of the *E. faecalis* polypeptides is synthesized and used. The amplified fragment is digested with the restriction endonucleases and then purified again on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC4 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene are used for transfection. Five  $\mu$ g of the expression plasmid pC4 is cotransfected with 0.5  $\mu$ g of the plasmid pSVneo using a lipid-mediated transfection agent such as Lipofectin™ or LipofectAMINE™ (LifeTechnologies Gaithersburg, MD). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml

flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1  $\mu$ M, 2  $\mu$ M, 5  $\mu$ M, 10  $\mu$ M, 20  $\mu$ M). The same procedure is repeated until clones are  
5 obtained which grow at a concentration of 100-200  $\mu$ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

### 9. Quantitative Murine Soft Tissue Infection Model for

#### 10 *E. faecalis*

Compositions of the present invention, including polypeptides and peptides, are assayed for their ability to function as vaccines or to enhance/stimulate an immune response to a bacterial species (e.g., *E. faecalis*) using the following quantitative murine soft tissue infection model. Mice (e.g., NIH Swiss female mice, approximately 7 weeks  
15 old) are first treated with a biologically protective effective amount, or immune enhancing/stimulating effective amount of a composition of the present invention using methods known in the art, such as those discussed above. See, e.g., Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988). An example of an appropriate starting dose is 20ug per animal.

20 The desired bacterial species used to challenge the mice, such as *E. faecalis*, is grown as an overnight culture. The culture is diluted to a concentration of  $5 \times 10^8$  cfu/ml, in an appropriate media, mixed well, serially diluted, and titered. The desired doses are further diluted 1:2 with sterilized Cytodex 3 microcarrier beads preswollen in sterile PBS (3g/100ml). Mice are anesthetized briefly until docile, but still mobile and  
25 injected with 0.2 ml of the Cytodex 3 bead/bacterial mixture into each animal subcutaneously in the inguinal region. After four days, counting the day of injection as day one, mice are sacrificed and the contents of the abscess is excised and placed in a 15 ml conical tube containing 1.0ml of sterile PBS. The contents of the abscess is then enzymatically treated and plated as follows.

30 The abscess is first disrupted by vortexing with sterilized glass beads placed in the tubes. 3.0mls of prepared enzyme mixture (1.0ml Collagenase D (4.0 mg/ml), 1.0ml Trypsin (6.0 mg/ml) and 8.0 mls PBS) is then added to each tube followed by a 20 min. incubation at 37C. The solution is then centrifuged and the supernatant drawn off. 0.5 ml dH2O is then added and the tubes are vortexed and then incubated for 10 min. at room  
35 temperature. 0.5 ml media is then added and samples are serially diluted and plated onto agar plates, and grown overnight at 37C. Plates with distinct and separate colonies are then counted, compared to positive and negative control samples, and quantified. The



method can be used to identify composition and determine appropriate and effective doses for humans and other animals by comparing the effective doses of compositions of the present invention with compositions known in the art to be effective in both mice and humans. Doses for the effective treatment of humans and other animals, using compositions of the present invention, are extrapolated using the data from the above experiments of mice. It is appreciated that further studies in humans and other animals may be needed to determine the most effective doses using methods of clinical practice known in the art.

#### 10. Murine Systemic Neutropenic Model for *E. faecalis* Infection

Compositions of the present invention, including polypeptides and peptides, are assayed for their ability to function as vaccines or to enhance/stimulate an immune response to a bacterial species (e.g., *E. faecalis*) using the following qualitative murine systemic neutropenic model. Mice (e.g., NIH Swiss female mice, approximately 7 weeks old) are first treated with a biologically protective effective amount, or immune enhancing/stimulating effective amount of a composition of the present invention using methods known in the art, such as those discussed above. See, e.g., Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988). An example of an appropriate starting dose is 20ug per animal.

Mice are then injected with 250 - 300 mg/kg cyclophosphamide intraperitoneally. Counting the day of C.P. injection as day one, the mice are left untreated for 5 days to begin recovery of PMNL'S.

The desired bacterial species used to challenge the mice, such as *E. faecalis*, is grown as an overnight culture. The culture is diluted to a concentration of  $5 \times 10^8$  cfu/ml, in an appropriate media, mixed well, serially diluted, and titered. The desired doses are further diluted 1:2 in 4% Brewer's yeast in media.

Mice are injected with the bacteria/brewer's yeast challenge intraperitoneally. The Brewer's yeast solution alone is used as a control. The mice are then monitored twice daily for the first week following challenge, and once a day for the next week to ascertain morbidity and mortality. Mice remaining at the end of the experiment are sacrificed.

The method can be used to identify compositions and determine appropriate and effective doses for humans and other animals by comparing the effective doses of compositions of the present invention with compositions known in the art to be effective in both mice and humans. Doses for the effective treatment of humans and other animals, using compositions of the present invention, are extrapolated using the data from the above experiments of mice. It is appreciated that further studies in humans

and other animals may be needed to determine the most effective doses using methods of clinical practice known in the art.

The disclosure of all publications (including patents, patent applications, journal articles, laboratory manuals, books, or other documents) cited herein are hereby  
5 incorporated by reference in their entireties.

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended as single illustrations of individual aspects of the invention. Functionally equivalent methods and components are within the scope of the invention, in addition to those shown and described herein and will become apparant to  
10 those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	Orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	KSP nt length
3	2	423	1226	gb U24692	"Enterococcus faecalis pyrimidine biosynthesis D (pyrD) gene, complete cds"	99	229
47	14	17085	16216	gb M81466	"Enterococcus faecalis RecA protein (recA) gene, partial cds"	98	308
52	1	50	1441	emb X62755 SFNPRG	S.faecalis npr gene for NADH peroxidase	98	1374
52	2	2456	1494	emb X62755 SFNPRG	S.faecalis npr gene for NADH peroxidase	100	209
61	1	2	358	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX)"	99	318
61	2	467	1975	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX)"	98	1297
61	3	1749	1967	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX)"	100	136
61	4	1990	2949	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-	100	960

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	Orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
					Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>"		
61	5	2112	2399	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>"	100	288
61	6	2922	3794	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>"	100	873
61	7	3671	4762	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>"	99	1092
61	8	4312	3860	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>"	100	453
61	9	4653	5783	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB),	100	1131

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
61	10	5750	6397	gb U35369	D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>"	99	648
61	11	7158	6784	gb U35369	"Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>"	100	161
67	1	3	809	gb U24692	"Enterococcus faecalis pyrimidine biosynthesis D (pyrD) gene, complete cds"	98	807
67	2	781	1512	gb U24692	"Enterococcus faecalis pyrimidine biosynthesis D (pyrD) gene, complete cds"	93	92
69	1	1	228	gb U60038	"Enterococcus faecalis major cold-shock protein (cspa) gene, partial cds"	100	136
72	15	15814	19737	emb X62656 EFASP1	"E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes"	92	2504
72	16	19739	20155	emb X62657 EFORF3	E.faecalis plasmid pAD1 DNA for orf3	96	341
75	1	3	365	emb Z19137 EFPTSHGN	E.faecalis of ptsH gene encoding HPr	100	267
83	12	8766	7432	emb X78425 EFPPB5	E.faecalis pbp5 gene	98	416
83	13	8869	9699	emb X78425 EFPPB5	E.faecalis pbp5 gene	99	819
83	14	9612	10913	emb X78425 EFPPB5	E.faecalis pbp5 gene	99	1203
83	15	10943	11746	emb X78425 EFPPB5	E.faecalis pbp5 gene	97	286

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	KSP nt length
84	2	1657	3558	emb X86176 EFRPODDNE	<i>E. faecalis</i> dnaE and rpoD gene	99	797
84	3	3649	4773	emb X86176 EFRPODDNE	<i>E. faecalis</i> dnaE and rpoD gene	99	1125
84	4	4913	7000	emb X86176 EFRPODDNE	<i>E. faecalis</i> dnaE and rpoD gene	99	301
104	2	4018	2900	gb U36195	"Enterococcus faecalis pyrAa gene, partial cds"	93	310
108	7	5875	5183	gb M58002	"Streptococcus faecalis bacterial cell wall hydrolase gene, complete cds"	98	252
145	8	8193	7234	gb U03756	"Enterococcus faecalis endocarditis specific antigen gene, complete cds"	99	960
145	9	8836	8147	gb U03756	"Enterococcus faecalis endocarditis specific antigen gene, complete cds"	100	132
147	3	2096	3418	emb X68847 SFNOXAA	<i>S. faecalis</i> nox gene for NADH oxidase	99	1301
154	4	2160	2492	emb X17092 PPRRA	Plasmid pAM-beta-1 (from <i>S. faecalis</i> ) replication region DNA	93	294
154	10	5935	6294	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	99	355
154	11	6279	6584	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	98	89
154	12	7882	7097	gb U86375	"Enterococcus faecalis ermB regulator and adenine methylase (ermB) genes, complete cds"	99	736
154	13	8750	8043	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	99	498
159	1	158	1483	gb M58002	"Streptococcus faecalis bacterial cell wall hydrolase gene, complete cds"	98	1323
159	2	807	157	gb M58002	"Streptococcus faecalis bacterial cell wall hydrolase gene, complete cds"	99	651
159	3	1395	2192	gb M58002	"Streptococcus faecalis bacterial cell wall hydrolase gene, complete cds"	93	350

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	Orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
216	2	282	1841	gb M90060	"Streptococcus faecalis H+ ATPase a (atpB), b (atpF), c (atpE), alpha (atpA), beta (atpD), gamma (atpG), delta (atpH), and epsilon (atpC) subunits, complete cds"	81	1558
216	4	2809	2967	gb M90060	"Streptococcus faecalis H+ ATPase a (atpB), b (atpF), c (atpE), alpha (atpA), beta (atpD), gamma (atpG), delta (atpH), and epsilon (atpC) subunits, complete cds"	86	132
216	5	2940	4244	gb M90060	"Streptococcus faecalis H+ ATPase a (atpB), b (atpF), c (atpE), alpha (atpA), beta (atpD), gamma (atpG), delta (atpH), and epsilon (atpC) subunits, complete cds"	83	1293
238	3	1814	2218	gb M38386	"Streptococcus faecalis mtlf enzymeIII, mannitol-mtld-phosphate- dehydrogenase"	96	302
238	4	2182	2670	gb M38386	"Streptococcus faecalis mtlf enzymeIII, mannitol-mtld-phosphate- dehydrogenase"	98	480
238	5	2634	3839	gb M38386	"Streptococcus faecalis mtlf enzymeIII, mannitol-mtld-phosphate- dehydrogenase"	96	459
261	2	1397	510	emb Z12296 EFSPREG	E.faecalis sprE gene for serine proteinase homologue	98	888
261	3	2474	1413	dbj D85393 ENEGE1E	"Enterococcus faecalis DNA for gelatinase, complete cds"	98	1051
261	4	2974	2417	dbj D85393 ENEGE1E	"Enterococcus faecalis DNA for gelatinase, complete cds"	97	516
275	3	1472	1044	gb L23802	"Enterococcus faecalis pore forming, cell wall enzyme, regulatory, and dehydroquinase homologue proteins (ebsA, ebsB, ebsC, and ebsD) genes, complete cds with repeat region"	98	422
275	4	1581	2018	gb L23802	"Enterococcus faecalis pore forming, cell wall enzyme, regulatory, and dehydroquinase homologue proteins"	97	438

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	Orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
275	5	2789	2148	gb L23802	(ebsA, ebsB, ebsC, and ebsD) genes, complete cds with repeat region"	98	642
275	6	3475	2660	gb L23802	"Enterococcus faecalis pore forming, cell wall enzyme, regulatory, and dehydroquinase homologue proteins (ebsA, ebsB, ebsC, and ebsD) genes, complete cds with repeat region"	98	790
287	2	1565	558	emb X17092 PPRRA	Plasmid pAM-beta-1 (from <i>S. faecalis</i> ) replication region DNA	97	991
287	3	2049	1582	emb X17092 PPRRA	Plasmid pAM-beta-1 (from <i>S. faecalis</i> ) replication region DNA	97	461
287	6	2639	3346	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	99	498
294	11	4519	4211	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	100	50
302	1	1	1755	emb X62658 EFSEA1	<i>E. faecalis</i> plasmid pAD1 seal gene and orf1	83	1755
302	2	2310	2687	emb X17214 SFPASA1	<i>S. faecalis</i> plasmid pAD1 asal gene for aggregation substance and ORF 1	100	378
302	3	2865	3329	emb X17214 SFPASA1	<i>S. faecalis</i> plasmid pAD1 asal gene for aggregation substance and ORF 1	99	463
316	4	2724	2110	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	100	248
346	5	2224	2880	emb X62755 SFNPRG	<i>S. faecalis</i> npr gene for NADH peroxidase	98	351



Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Indent	HSP nt length
349	2	686	907	dbj D78257 D78257	"Enterococcus faecalis plasmid pY117 genes for BacA, BacB, ORF3, ORF4, ORF5, ORF6, ORF7, ORF8, ORF9, ORF10, ORF11, partial cds."	83	200
355	1	3	1166	emb X17214 SFPASAL	S. faecalis plasmid pAD1 asal gene for aggregation substance and ORF 1	97	1100
355	2	1102	1548	emb X17214 SFPASAL	S. faecalis plasmid pAD1 asal gene for aggregation substance and ORF 1	94	432
355	3	1663	2037	emb X62657 EFORF3	E. faecalis plasmid pAD1 DNA for orf3	99	337
355	4	2035	2445	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	99	411
355	5	2558	2851	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	96	280
355	6	2838	3299	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	97	430
355	7	3236	3739	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	97	279
355	8	3696	4529	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	97	537
355	9	4587	5870	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	98	718
355	10	5843	6490	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	99	224
355	11	6471	6890	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	96	361
355	12	6881	7204	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	98	324
355	13	7191	8231	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	98	984
355	14	8218	8496	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	99	279
355	15	8412	8885	emb X96977 EFPAD1ORF	"E. faecalis plasmid pAD1, open reading frames"	100	474

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	Orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
					frames"		
355	17	9479	9952	emb X96977 EFPAD1ORF	"E.faecalis plasmid PAD1, open reading frames"	98	417
365	1	3	380	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	100	248
370	1	1	1299	dbj D78016 ENEPPD1A	"Enterococcus faecalis Plasmid pPD1 genes for REPB, REPA, TRAC, TRAB, TRAA, iPD1, TRAE, TRAF, complete cds and partial cds"	73	1267
407	3	963	2162	gb U38590	"Enterococcus faecalis plasmid pCF10 PrGN, PrGO, and PrGP genes, complete cds"	98	257
407	5	3811	4131	gb U38590	"Enterococcus faecalis plasmid pCF10 PrGN, PrGO, and PrGP genes, complete cds"	86	317
417	1	42	419	gb U00681	"Enterococcus faecalis plasmid PAD1 Trab (trab) gene, complete cds (trac) and (repA) genes, partial cds"	98	304
417	2	313	41	gb U00681	"Enterococcus faecalis plasmid PAD1 Trab (trab) gene, complete cds (trac) and (repA) genes, partial cds"	97	198
417	3	440	754	gb U00681	"Enterococcus faecalis plasmid PAD1 Trab (trab) gene, complete cds (trac) and (repA) genes, partial cds"	100	219
426	1	112	462	emb Z49243 EF4110SOD	E.faecalis partial sod gene for superoxide dismutase (strain=BM4110)	98	291
426	2	628	419	emb Z49243 EF4110SOD	E.faecalis partial sod gene for superoxide dismutase (strain=BM4110)	100	148
426	3	456	725	emb Z49243 EF4110SOD	E.faecalis partial sod gene for superoxide dismutase (strain=BM4110)	100	148
429	1	840	79	emb X62658 EFSEA1	E.faecalis plasmid PAD1 seal gene and orfy	98	737
429	2	1087	767	emb X62658 EFSEA1	E.faecalis plasmid PAD1 seal gene and orfy	99	321
429	4	2765	2460	gb U17153	"Enterococcus faecalis plasmid pJh1	98	89

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	Orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
					tetracycline resistant (tetL) gene, complete cds		
429	5	3166	2750	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	99	413
435	5	2731	2324	gb M38052	"Enterococcus faecalis cytolysin B transport protein gene, complete cds"	97	97
459	2	1330	1067	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	99	248
506	1	1242	4	emb X17214 SFPASA1	S. faecalis plasmid pAD1 asal gene for aggregation substance and ORF 1	99	1144
514	3	1496	1113	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	100	248
527	2	1733	1371	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	98	153
544	1	309	4	gb U38590	"Enterococcus faecalis plasmid pCF10 PrgN, PrgO, and PrgP genes, complete cds"	95	306
561	1	3	761	dbj D78016 ENEPPD1A	"Enterococcus faecalis plasmid pPD1 genes for REPB, REPA, TRAC, TRAB, TRAA, iPD1, TRAE, TRAF, complete cds and partial cds"	77	528
561	2	772	1566	gb U00601	"Enterococcus faecalis plasmid pAD1 Trab (trab) gene, complete cds (trac) and (repA) genes, partial cds"	99	795
566	3	874	2037	dbj D78016 ENEPPD1A	"Enterococcus faecalis plasmid pPD1 genes for REPB, REPA, TRAC, TRAB, TRAA, iPD1, TRAE, TRAF, complete cds and partial cds"	90	1160
581	1	398	3	emb X96977 EPPAD1ORF	"E. faecalis plasmid pAD1, open reading	100	393

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	Orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
581	2	908	540	emb X96977 EFPAD1ORF	frames" "E.faecalis plasmid pAD1, open reading frames"	100	369
597	1	573	7	gb M38052	"Enterococcus faecalis cytolsin B transport protein gene, complete cds"	99	566
597	2	1247	516	gb M38052	"Enterococcus faecalis cytolsin B transport protein gene, complete cds"	97	701
604	7	3265	2903	gb U17153	"Enterococcus faecalis plasmid pJh1 tetracycline resistant (tetL) gene, complete cds"	100	143
618	1	1	534	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	99	470
622	1	864	16	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	99	849
622	2	1317	862	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	99	256
622	3	1586	1311	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	99	248
624	6	5641	8001	gb U56286	"Enterococcus faecalis gyrase A (gyrA) gene, partial cds"	98	219
635	1	516	953	dbj D78257 D78257	"Enterococcus faecalis plasmid pY17 genes for BacA, BacB, ORF3, ORF4, ORF5, ORF6, ORF7, ORF8, ORF9, ORF10, ORF11, partial cds"	94	404

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
635	2	920	1222	dbj D78257 D78257	"Enterococcus faecalis plasmid pY117 genes for BacA, BacB, ORF3, ORF4, ORF5, ORF6, ORF7, ORF8, ORF9, ORF10, ORF11, partial cds"	83	299
637	1	3	545	emb X62656 EFASP1	"E. faecalis plasmid pPD1 aspl and URFs pd57, pd125 and pd113 genes"	92	506
658	2	1198	365	gb M38052	"Enterococcus faecalis cytolysin B transport protein gene, complete cds"	100	819
658	3	1446	1189	gb M38052	"Enterococcus faecalis cytolysin B transport protein gene, complete cds"	98	258
664	1	490	65	emb X62658 EFSEA1	E. faecalis plasmid pAD1 seal gene and orfy	88	423
664	2	737	417	emb X62658 EFSEA1	E. faecalis plasmid pAD1 seal gene and orfy	94	321
743	1	561	4	dbj D78016 ENEPPDIA	"Enterococcus faecalis Plasmid pPD1 genes for REPB, REPA, TRAC, TRAB, TRAA, iPD1, TRAE, TRAF, complete cds and partial cds"	87	305
747	2	1139	324	gb M38052	"Enterococcus faecalis cytolysin B transport protein gene, complete cds"	99	691
747	3	577	783	gb M38052	"Enterococcus faecalis cytolysin B transport protein gene, complete cds"	100	207
747	4	1474	1133	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	99	248
777	1	401	3	gb M38052	"Enterococcus faecalis cytolysin B transport protein gene, complete cds"	100	335
816	1	793	512	gb M13771	"Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds"	100	243
842	1	418	89	emb X17214 SFPASA1	S. faecalis plasmid pAD1 asal gene for aggregation substance and ORF 1	91	303
842	2	856	605	emb X62658 EFSEA1	E. faecalis plasmid pAD1 seal gene and orfy	92	246

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	Orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
847	1	1481	3	emb X62658 EFSEA1	E.faecalis plasmid pAD1 seal gene and orfY	92	1479
864	1	36	1106	emb X62658 EFSEA1	E.faecalis plasmid pAD1 seal gene and orfY	93	945
864	2	1571	3550	emb X62656 EFASPI	"E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes"	96	1979
872	1	263	3	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	98	261
874	1	833	693	dbj D31675 ENE16RNA8	"Enterococcus faecalis 16S ribosomal RNA, partial sequence"	100	98
878	1	302	30	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	94	94
878	2	263	445	gb U17153	"Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds"	99	181
921	1	748	26	emb X62658 EFSEA1	E.faecalis plasmid pAD1 seal gene and orfY	95	612
929	1	484	2	emb X62658 EFSEA1	E.faecalis plasmid pAD1 seal gene and orfY	99	409
946	1	3	422	emb X62657 EFORF3	E.faecalis plasmid pAD1 DNA for orf3	99	341
946	2	420	830	emb X96977 EFPAD1ORF	"E.faecalis plasmid pAD1, open reading frames"	98	411
946	3	866	1123	emb X96977 EFPAD1ORF	"E.faecalis plasmid pAD1, open reading frames"	96	230
947	1	112	498	emb X62656 EFASPI	"E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes"	96	378
951	1	484	26	emb X62658 EFSEA1	E.faecalis plasmid pAD1 seal gene and orfY	95	353
956	1	3	545	emb X62656 EFASPI	"E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes"	96	543
956	2	524	721	emb X62656 EFASPI	"E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes"	94	161
957	1	616	2	emb X96977 EFPAD1ORF	"E.faecalis plasmid pAD1, open reading frames"	99	615
957	2	42	686	emb X96977 EFPAD1ORF	"E.faecalis plasmid pAD1, open reading frames"	99	595

Table 1: *E. faecalis* - Coding regions containing known sequences

Contig ID	Orf ID	Start (nt)	Stop (nt)	Match Accession	Match Gene Name	Percent Ident	HSP nt length
					frames"		
968	1	1	456	emb X62656 EFASP1	"E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes"	96	366
968	2	339	641	emb X62656 EFASP1	"E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes"	95	158
968	3	395	658	emb X62656 EFASP1	"E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes"	94	126
977	1	5	943	emb X17214 SFPASA1	S. faecalis plasmid pAD1 asal gene for aggregation substance and ORF 1	99	847
982	1	376	2	emb X62658 EFSEA1	E.faecalis plasmid pAD1 seal gene and orfy	95	365
985	1	85	471	emb X62656 EFASP1	"E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes"	91	362

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
137	3	3208	2003	gi 152947	transposase [Staphylococcus aureus]	100	100
154	14	9166	9750	gi 141861	traA gene product [Plasmid pAD1]	100	100
276	16	11268	11047	gnl PID e284733	C34B7.2 [Caenorhabditis elegans]	100	71
287	1	485	234	gi 152947	transposase [Staphylococcus aureus]	100	100
287	7	3454	3765	gi 152947	transposase [Staphylococcus aureus]	100	100
292	6	3001	4185	gi 488330	alpha-amylase [unidentified cloning vector]	100	100
429	3	2013	1654	gi 141863	regulatory protein [Plasmid pAD1]	100	100
604	3	1243	1043	gi 559860	clyLs [Plasmid pAD1]	100	98
604	4	1492	1268	gi 559859	clyL1 [Plasmid pAD1]	100	100
656	7	7592	6834	gi 488339	alpha-amylase [unidentified cloning vector]	100	100
658	1	312	4	gi 152947	transposase [Staphylococcus aureus]	100	100
674	3	1236	1589	gi 1196996	unknown protein [Transposon Tn10]	100	98
700	1	375	4	gi 152947	transposase [Staphylococcus aureus]	100	100
961	1	1	450	gi 152947	transposase [Staphylococcus aureus]	100	100
72	17	20153	21040	gi 150556	surface protein [Plasmid pCF10]	99	99
99	5	3117	1933	gi 1006839	malic enzyme [Streptococcus bovis]	99	99
154	3	1955	1491	gi 149482	transposase [Lactococcus lactis]	99	99
326	3	3030	1774	pir S16989 S16989	dihydrolipoamide S-acetyltransferase (EC 2.3.1.12) - Enterococcus faecalis	99	98
407	6	4636	4235	gi 141859	replication-associated protein [Plasmid pAD1]	99	99
692	1	3	485	gi 559861	clyM [Plasmid pAD1]	99	99
99	6	3904	3134	gi 1146122	L-malate permease [Streptococcus bovis]	98	98
326	4	3358	3002	pir S16989 S16989	dihydrolipoamide S-acetyltransferase (EC 2.3.1.12) - Enterococcus faecalis	98	97
346	1	606	4	gi 1146122	L-malate permease [Streptococcus bovis]	98	98
367	31	14415	13999	gi 1644226	ribosomal protein S10 [Bacillus subtilis]	98	88
367	6	2797	2495	gi 142459	initiation factor 1 [Bacillus subtilis]	97	88
407	9	5454	4894	gi 141858	replication-associated protein [Plasmid pAD1]	97	97



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
					pAD1		
497	6	3514	3762	gi 532552	ORF19 [Enterococcus faecalis]	97	87
558	1	1	399	gi 46638	ORF 2 (AA 1 - 236) [Staphylococcus aureus]	97	97
829	1	169	2	gnl PID e283110	femD [Staphylococcus aureus]	97	86
407	8	4970	4599	gi 141858	replication-associated protein [Plasmid pAD1]	96	96
777	2	1102	380	gi 559861	clyM [Plasmid pAD1]	96	96
23	33	20797	21126	gnl PID e223402	DNA topoisomerase IV C subunit [Streptococcus pneumoniae]	95	80
32	5	3454	3071	gi 147194	phnA protein [Escherichia coli]	95	87
95	8	5493	6875	gi 391682	Na <sup>+</sup> -ATPase beta subunit [Enterococcus hirae]	95	89
138	25	16587	16745	gi 143136	L-lactate dehydrogenase [Bacillus megaterium]	95	70
367	20	9198	8797	gi 40150	L14 protein (AA 1-122) [Bacillus subtilis]	95	90
367	21	9519	9223	gi 1044973	ribosomal protein L17 [Bacillus subtilis]	95	89
439	2	846	1241	gi 488334	alpha-amylase [unidentified cloning vector]	95	94
604	1	792	4	gi 559861	clyM [Plasmid pAD1]	95	93
722	1	1	504	gi 47453	ribosomal protein S12 [Streptococcus pneumoniae]	95	94
17	8	7317	7676	gi 532554	ORF21 [Enterococcus faecalis]	94	86
95	2	1288	1791	gi 416405	Na <sup>+</sup> -ATPase K subunit [Enterococcus hirae]	94	88
97	3	2481	1432	gi 1750264	heat shock protein 70 [Streptococcus pneumoniae]	94	90
117	5	2700	3842	gi 467376	unknown [Bacillus subtilis]	94	89
327	3	3283	3762	gi 153566	ORF (19K protein) [Enterococcus faecalis]	94	87
327	5	4782	5054	gi 153568	H <sup>+</sup> ATPase [Enterococcus faecalis]	94	82
387	4	3608	1728	gi 153661	translational initiation factor IF2 [Enterococcus faecium] sp P18311 IF2_ENTFC INITIATION FACTOR IF-2.	94	88

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
455	1	2	259	gi 532549	ORF16 [Enterococcus faecalis]	94	82
97	2	1444	677	gi 450684	dnak gene product [Lactococcus lactis]	93	83
188	2	1690	1911	gi 43865	nifJ gene product [Klebsiella pneumoniae]	93	78
216	6	4234	4680	gi 153574	H+ ATPase [Enterococcus faecalis]	93	86
298	2	2798	1221	gi 143012	GMP synthetase [Bacillus subtilis]	93	86
329	2	1538	771	gi 153826	adhesin B [Streptococcus sanguis]	93	83
367	15	7675	7247	gi 1044978	ribosomal protein S8 [Bacillus subtilis]	93	82
722	2	527	1030	gi 1644222	ribosomal protein S7 [Bacillus subtilis]	93	83
803	1	657	151	gi 1196998	unknown protein [Transposon Tn10]	93	93
962	1	130	636	gi 152947	transposase [Staphylococcus aureus]	93	92
237	12	6056	6385	gi 963038	ArpU [Enterococcus hirae]	92	76
309	4	8218	4541	gi 402363	RNA polymerase beta-subunit [Bacillus subtilis] sp P37870 RPOB_BACSU DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC .7.7.6) (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT).	92	82
329	4	2529	1717	gi 310632	hydrophobic membrane protein [Streptococcus gordonii] sp P42361 P29K_STRGC 29 KD MEMBRANE PROTEIN IN PSAA 5' REGION ORF1.	92	78
367	4	1942	1544	gi 142462	ribosomal protein S11 [Bacillus subtilis]	92	82
367	8	3648	3457	pir C44859 C44859	adenylate kinase - Bacillus sp. (fragment)	92	88
367	12	6183	5641	gi 1044981	ribosomal protein S5 [Bacillus subtilis]	92	81
367	17	8427	7885	pir A29102 R5BS5F	ribosomal protein L5 - Bacillus stearothermophilus	92	83
527	1	1404	373	gi 153092	replication protein [Staphylococcus aureus]	92	81
701	1	2	352	gi 143793	tyrosyl-tRNA synthetase [Bacillus caldotenax]	92	74
23	28	17420	17566	sp P45692 EUTX_SAL TY	ETHANOLAMINE UTILIZATION PROTEIN EUTX (FRAGMENT).	91	73

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
57	5	4129	4701	gi 1595810	type-I signal peptidase SpsB [Staphylococcus aureus]	91	67
57	12	13281	13970	gnl PID e254999	phenylalanyl-tRNA synthetase beta subunit [Bacillus subtilis]	91	75
156	5	4609	6474	gi 1303804	YgeQ [Bacillus subtilis]	91	79
216	3	1848	2765	gi 153572	H+ ATPase [Enterococcus faecalis]	91	81
367	24	10802	10128	gi 1165309	S3 [Bacillus subtilis]	91	78
415	1	452	883	pir B56272 B56272	probable pheromone-responsive regulatory protein R - Enterococcus faecalis plasmid pCF10	91	90
466	2	1313	2065	gi 142443	adenylosuccinate synthetase [Bacillus subtilis] sp P29726 PURA_BACSU ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) IMP--ASPARTATE LIGASE).	91	79
545	1	1	345	gi 532549	ORF16 [Enterococcus faecalis]	91	80
572	1	8	652	gi 347998	uracil phosphoribosyltransferase [Streptococcus salivarius] sp P36399 UPP_STRSL PROBABLE URACIL PHOSPHORIBOSYLTRANSFERASE (EC 4.2.9) (UMP PYROPHOSPHORYLASE) (UPRTASE)..	91	78
599	1	8	343	gi 42029	ORF1 gene product [Escherichia coli]	91	75
600	2	585	779	pir B48396 B48396	ribosomal protein L33 - Bacillus stearothermophilus	91	81
652	1	394	2	gi 535662	transposase [Insertion sequence IS1251]	91	81
1	4	3465	2557	gi 1644224	elongation factor Tu [Bacillus subtilis]	90	83
17	19	14844	17297	gi 532549	ORF16 [Enterococcus faecalis]	90	77
52	3	2650	2811	gi 473902	alpha-acetolactate synthase [Lactococcus lactis]	90	68
74	9	5870	5469	gi 1653508	hypothetical protein [Synecocystis sp.]	90	52
75	3	1177	2091	gi 153615	phosphoenolpyruvate:sugar phosphotransferase system enzyme I Streptococcus salivarius]	90	83

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
117	10	6591	8126	gi 924848	inosine monophosphate dehydrogenase [Streptococcus pyogenes] pir JC4372 JC4372 IMP dehydrogenase (EC 1.1.1.205) - Streptococcus yogenes	90	80
276	1	577	95	gi 530798	LysB [Bacteriophage phi-LC3]	90	72
287	5	2611	2441	gi 1333835	copS gene product [Streptococcus pyogenes]	90	78
290	1	1	708	gi 897795	30S ribosomal protein [Pediococcus acidilactici] sp P49668 RS2_PEDAC 30S RIBOSOMAL PROTEIN S2.	90	75
309	3	4401	1093	gnl PID e187579	DNA-directed RNA polymerase [Listeria innocua]	90	81
367	22	9731	9513	pir A02825 R5B529	ribosomal protein L29 - Bacillus stearothermophilus	90	76
452	4	2224	2508	gi 434759	ORF [Homo sapiens]	90	54
455	2	2776	323	gi 532549	ORF16 [Enterococcus faecalis]	90	77
623	1	3	221	gi 460259	enolase [Bacillus subtilis]	90	80
624	5	3612	5615	gnl PID e208213	DNA gyrase [Streptococcus pneumoniae]	90	81
853	2	752	282	gnl PID e13389	translation initiation factor IF3 (AA 1-172) [Bacillus stearothermophilus]	90	82
966	1	1	462	gi 532549	ORF16 [Enterococcus faecalis]	90	83
1	3	2596	2219	gi 1661195	elongation factor-Tu [Streptococcus mutans]	89	78
1	5	4314	3556	gi 1644223	elongation factor G [Bacillus subtilis]	89	79
23	21	13990	14295	gi 466518	pduA [Salmonella typhimurium]	89	75
23	32	19927	20799	gnl PID e208211	DNA topoisomerase IV [Streptococcus pneumoniae]	89	83
42	2	349	1989	gi 287871	groEL gene product [Lactococcus lactis]	89	79
45	15	11835	12167	gi 150554	surface exclusion protein [Plasmid pCF10]	89	68
53	2	685	1797	gnl PID e221213	ClpX protein [Bacillus subtilis]	89	81
86	4	3374	4024	gi 537286	triosephosphate isomerase [Lactococcus lactis]	89	78

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
95	7	3677	5506	gi 912449	Na <sup>+</sup> -ATPase alpha subunit [Enterococcus hirae]	89	80
128	18	11348	11013	gi 466473	cellobiose phosphotransferase enzyme II' [Bacillus tearothermophilus]	89	60
132	1	180	2180	gi 153854	uvs402 protein [Streptococcus pneumoniae]	89	78
342	1	783	4	gi 1041115	TRAC [Plasmid pPD1]	89	79
367	23	10146	9691	sp P14577 RL16_BAC SU	50S RIBOSOMAL PROTEIN L16.	89	80
367	27	12377	11541	gi 1165306	L2 [Bacillus subtilis]	89	79
435	4	2424	2215	gi 559863	clyA [Plasmid pAD1]	89	89
466	3	1972	2736	gi 467328	adenylosuccinate synthetase [Bacillus subtilis]	89	75
512	3	999	1607	gi 1477776	ClpP [Bacillus subtilis]	89	73
518	1	1	174	gi 786163	Ribosomal Protein L10 [Bacillus subtilis]	89	76
604	2	1000	713	gi 559861	clyM [Plasmid pAD1]	89	89
615	2	888	691	gi 467469	unknown [Bacillus subtilis]	89	75
677	2	992	429	gi 1389732	S-adenosylmethionine synthetase [Bacillus subtilis]	89	76
677	3	1315	950	gi 1020317	S-adenosylmethionine synthetase [Staphylococcus aureus]	89	73
722	3	1102	1278	pir PW0010 PW0010	translation elongation factor G - Bacillus stearothermophilus (fragment)	89	72
850	1	464	3	gi 142521	deoxyribodipyrimidine photolyase [Bacillus subtilis] gnl PID e255102	89	72
					deoxyribodipyrimidine photolyase [Bacillus ubtilis]		
17	5	3711	4751	gi 532554	ORF21 [Enterococcus faecalis]	88	72
37	5	3322	3717	gi 1216488	uncharacterized open reading frame; hypothetical protein displaying similarity to a Bacillus subtilis hypothetical protein (Ylm [Streptococcus mutans])	88	75
39	6	2454	2630	sp P49865 NTPR_ENT	NTPR PROTEIN (FRAGMENT).	88	77

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
				HR			
48	3	1740	2666	gi 557492	dihydroxynaphthoic acid (DHNA) synthetase [Bacillus subtilis] gi 143186	88	75
					dihydroxynaphthoic acid (DHNA) synthetase [Bacillus ubtilis]		
63	5	2753	3607	gi 1064814	homologous to sp:PHOP_BACSUB [Bacillus subtilis]	88	77
86	2	1004	2047	gi 153763	plasmin receptor [Streptococcus pyogenes]	88	79
104	6	6431	6213	gi 431231	uracil permease [Bacillus caldolyticus]	88	60
110	19	18174	16891	gi 217040	acid glycoprotein [Streptococcus pyogenes]	88	72
145	10	9040	8834	gi 393268	29-kiloDalton protein [Streptococcus pneumoniae] sp P42362 P29K_STRPN 29 KD MEMBRANE PROTEIN IN PSAA 5'REGION ORF1.	88	71
151	1	1620	316	gi 143366	adenylosuccinate lyase (PUR-B) [Bacillus subtilis] pir C29326 WZESDS	88	78
					adenylosuccinate lyase (EC 4.3.2.2) - Bacillus ubtilis		
171	10	9676	10119	gi 1591672	phosphate transport system ATP-binding protein [Methanococcus jannaschii]	88	63
190	3	1997	975	gi 532554	ORF21 [Enterococcus faecalis]	88	76
229	6	5712	5954	gi 143648	ribosomal protein L28 [Bacillus subtilis]	88	70
270	2	895	1869	gi 1303828	YqfJ [Bacillus subtilis]	88	75
275	7	3761	3552	gi 425474	SMDR1 [Schistosoma mansoni]	88	72
293	1	614	3	gi 1783246	highly homologous to many ATP-binding transport proteins; hypothetical [Bacillus subtilis]	88	80
367	1	485	72	gi 142464	ribosomal protein L17 [Bacillus subtilis]	88	76
367	5	2335	1961	gi 1044989	ribosomal protein S13 [Bacillus subtilis]	88	80
367	16	7887	7681	pir S48688 S48688	ribosomal protein S14 - Bacillus stearothermophilus	88	83
598	1	1006	23	gi 565287	transposase-like protein of PS3IS [thermophilic bacterium PS3]	88	66

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					pir JC4292 JC4292 insertion sequence element 1341 - thermophilic acterium PS-3		
600	3	1640	882	gi 763052	integrase [Bacteriophage T270]	88	68
669	1	2	514	gi 153801	enzyme scr-II [Streptococcus mutans]	88	75
808	2	624	394	gi 1574781	exodeoxyribonuclease V (recB) [Haemophilus influenzae]	88	77
871	1	714	229	gi 1574120	branched-chain-amino-acid transaminase [Haemophilus influenzae]	88	79
979	1	1	384	gnl PID e187579	DNA-directed RNA polymerase [Listeria innocua]	88	78
983	1	34	282	gi 40026	homologous to E.coli gida [Bacillus subtilis]	88	78
47	5	6799	5810	gi 532204	prs [Listeria monocytogenes]	87	79
69	3	2033	750	gi 1377831	unknown [Bacillus subtilis]	87	74
73	2	1432	167	gi 143434	Rho Factor [Bacillus subtilis]	87	76
76	5	2412	3740	gi 496283	lysin [Bacteriophage Tuc2009]	87	75
88	3	1600	2016	gnl PID e137596	heat shock induced protein HtpO [Lactobacillus leichmannii]	87	75
89	7	6003	5608	gi 1695686	pyruvate carboxylase [Bacillus stearothermophilus]	87	77
93	1	283	119	gi 1124825	unknown protein [Chlamydia trachomatis]	87	56
104	1	2945	3	gnl PID e199387	carbamoyl-phosphate synthase [Lactobacillus plantarum]	87	75
124	4	3191	2274	gi 995767	UDP-glucose pyrophosphorylase [Streptococcus pyogenes]	87	76
273	2	608	1108	gi 1184680	polynucleotide phosphorylase [Bacillus subtilis]	87	76
293	2	1020	532	gi 153741	ATP-binding protein [Streptococcus mutans]	87	74
326	5	4534	3533	gi 143378	pyruvate decarboxylase (E-1) beta subunit [Bacillus subtilis] gi 1377836 pyruvate decarboxylase E-1 beta subunit [Bacillus ubtilis]	87	74

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
334	3	3182	3340	pir A36324 A36324	growth arrest-specific protein - mouse	87	50
337	1	1382	186	gi 308861	GTG start codon [Lactococcus lactis]	87	75
338	8	6925	5723	gi 149575	L(+)-lactate dehydrogenase [Lactobacillus casei] sp P00343 LDH_LACCA L-LACTATE DEHYDROGENASE (EC 1.1.1.27). (SUB -326)	87	73
367	18	8782	8450	pir A02819 RSBS24	ribosomal protein L24 - Bacillus stearothermophilus	87	70
388	2	410	183	gnl PID e225674	unknown [Schizosaccharomyces pombe]	87	75
440	1	466	1797	gi 520754	putative [Bacillus subtilis]	87	75
508	1	694	137	gi 496558	orfX [Bacillus subtilis]	87	73
654	3	530	802	pir A47079 A47079	heat shock protein DnaJ - Lactococcus lactis	87	70
18	1	3	413	gi 46912	ribosomal protein L13 [Staphylococcus carnosus]	86	70
18	2	406	819	pir S08564 R3BS9	ribosomal protein S9 - Bacillus stearothermophilus	86	73
50	1	84	1148	gi 452398	threonine synthase [Bacillus sp.]	86	74
74	14	10547	10080	gi 1314299	ORF6; putative glutamyl-tRNA-transferase; similar to glutamyl-tRNA-transferase from Bacillus subtilis [Listeria monocytogenes]	86	74
95	5	3176	3406	gi 487276	Na+ -ATPase subunit C [Enterococcus hirae]	86	62
114	8	9216	10313	gi 853776	peptide chain release factor 1 [Bacillus subtilis] pir S55437 S55437 peptide chain release factor 1 - Bacillus ubtilis	86	69
115	2	501	899	gi 551879	ORF 1 [Lactococcus lactis]	86	70
164	26	25639	25842	pir S34762 S34762	L-serine dehydratase beta chain - Clostridium sp.	86	81
243	2	2143	1082	gi 143607	sporulation protein [Bacillus subtilis]	86	70
255	1	2	196	gi 755604	unknown [Bacillus subtilis]	86	64
257	3	3565	983	gi 928832	ORF259; putative [Lactococcus lactis] phage BK5-T	86	66



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
273	3	943	1314	gi 1184680	polynucleotide phosphorylase [Bacillus subtilis]	86	65
288	2	554	1087	gi 153033	tagatose 6-phosphate isomerase [Staphylococcus aureus] pir B38158 B38158 galactose-6-phosphate isomerase 19K chain - taphylococcus aureus	86	74
327	7	5183	5722	gi 153569	H+ ATPase [Enterococcus faecalis]	86	71
345	7	5111	5620	gi 1314294	ORF1; putative 17 kDa protein [Listeria monocytogenes]	86	63
350	3	1900	2781	gi 511015	dihydroorotate dehydrogenase A [Lactococcus lactis] sp P54321 PYDA_LACLC DIHYDROOROTATE DEHYDROGENASE A (EC 1.3.3.1) DIHYDROOROTATE OXIDASE A) (DHODHASE A).	86	73
363	3	3328	4233	gi 1657517	hypothetical protein [Escherichia coli]	86	59
367	25	11216	10851	gi 1165308	L22 [Bacillus subtilis]	86	68
367	26	11534	11220	gi 1165307	S19 [Bacillus subtilis]	86	77
367	30	13995	13453	gi 1165303	L3 [Bacillus subtilis]	86	75
393	1	1	660	sp P33898 G3P3_ECO LI	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE C (EC 1.2.1.12) (GAPDH-C).	86	77
396	1	1	192	gi 944942	RipX [Bacillus subtilis]	86	77
438	3	1279	1560	gi 1001878	CspL protein [Listeria monocytogenes]	86	75
510	1	1008	199	gi 473795	'ORF' [Escherichia coli]	86	71
510	2	1912	962	gi 473794	'ORF' [Escherichia coli]	86	76
539	1	705	4	gi 467477	unknown [Bacillus subtilis]	86	79
570	2	2069	1023	gi 881511	Ccpa protein [Lactobacillus casei]	86	72
654	2	240	575	pir A47079 A47079	heat shock protein DnaJ - Lactococcus lactis	86	77
677	1	431	102	gi 1389732	S-adenosylmethionine synthetase [Bacillus subtilis]	86	80
984	1	1	147	pir A56922 A56922	transcription factor shn - fruit fly (Drosophila melanogaster)	86	73

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
5	11	7720	8487	gi 41015	aspartate-tRNA ligase [Escherichia coli]	85	71
34	2	2133	1711	gi 47828	pyruvate kinase [Bacillus stearothermophilus]	85	75
97	4	2666	2517	pir S39341 S39341	grpE protein - Lactococcus lactis	85	66
103	2	1263	946	gi 143364	phosphoribosyl aminoimidazole carboxylase I (PUR-E) [Bacillus subtilis]	85	68
103	3	1465	1169	gi 143364	phosphoribosyl aminoimidazole carboxylase I (PUR-E) [Bacillus subtilis]	85	67
129	3	2395	3258	gi 143766	(thrSv) (EC 6.1.1.3) [Bacillus subtilis]	85	67
129	4	3240	4445	gi 143766	(thrSv) (EC 6.1.1.3) [Bacillus subtilis]	85	78
188	1	86	1447	gnl PID e214721	glutamine synthetase [Staphylococcus aureus]	85	71
217	3	673	1086	gi 520540	unknown [Bacillus subtilis]	85	72
241	2	1715	1086	gi 495089	recombinase [Staphylococcus aureus]	85	68
285	2	712	993	gi 40014	pot. ORF 446 (aa 1-446) [Bacillus subtilis]	85	77
293	3	1149	1595	gi 755604	unknown [Bacillus subtilis]	85	66
300	2	2738	2220	gi 289261	comE ORF2 [Bacillus subtilis]	85	72
305	2	1853	2695	pir S09411 S09411	spoIIIE protein - Bacillus subtilis	85	70
322	1	1	171	gi 153562	aspartate beta-semialdehyde dehydrogenase (EC 1.2.1.11) Streptococcus mutans]	85	67
327	4	4056	4784	gi 153567	H+ ATPase [Enterococcus faecalis]	85	66
367	10	5417	4959	pir A02795 R5BS15	ribosomal protein L15 - Bacillus stearothermophilus	85	76
383	3	3168	2953	gnl PID e274577	csp [Lactobacillus plantarum]	85	79
404	3	3069	2101	gi 143402	recombination protein (ttg start codon) [Bacillus subtilis] gi 1303923 RecN [Bacillus subtilis]	85	72
469	1	2	724	gi 508979	GTP-binding protein [Bacillus subtilis]	85	78
488	1	1	996	gi 532548	ORF15 [Enterococcus faecalis]	85	67
535	5	6468	4849	gi 634107	kdpB [Escherichia coli]	85	68

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
584	3	732	562	gi 467374	single strand DNA binding protein [Bacillus subtilis] sp P37455 SSB_BACSU SINGLE-STRAND BINDING PROTEIN (SSB) HELIX-DESTABILIZING PROTEIN).	85	75
695	1	78	500	gi 499384	orf189 [Bacillus subtilis]	85	75
836	1	1	357	gi 153801	enzyme scr-II [Streptococcus mutans]	85	69
17	20	17212	18813	gi 532548	ORF15 [Enterococcus faecalis]	84	68
23	31	18728	19987	gnl PID e208211	DNA topoisomerase IV [Streptococcus pneumoniae]	84	68
34	3	3112	2144	gi 143312	6-phospho-1-fructokinase (gtg start codon; EC 2.7.1.11) [Bacillus tearothermophilus]	84	69
36	1	1	1152	gi 1644223	elongation factor G [Bacillus subtilis]	84	73
49	12	6730	8190	gi 456319	74kDa protein [Bacteriophage FC1]	84	65
51	2	1379	1663	gi 468207	Submitter comments: A Mg2+ transporting P-type ATPase highly homologous with mgTB ATPase at 80 min on Salmonella chromosome. mediates the influx of Mg2+ only. Transcription regulated by xtracellular Mg2+ [Salmonella typhimurium]	84	71
95	6	3330	3707	gi 487277	Na+ -ATPase subunit G [Enterococcus hirae]	84	64
104	5	6250	5459	gnl PID e199440	aspartate carbamoyltransferase, aspartate transcarbamylase, carbamylaspartotranskinase [Lactobacillus plantarum]	84	65
105	6	4605	5273	gi 467411	recombination protein [Bacillus subtilis]	84	65
114	11	12278	12997	gi 556886	serine hydroxymethyltransferase [Bacillus subtilis] pir S49363 S49363 serine hydroxymethyltransferase - Bacillus subtilis	84	74
117	2	705	1484	gi 580906	B.subtilis genes rpmH, rnpA, 50kd, gidA and gidB [Bacillus subtilis] gi 467381 regulation of Spo0J and Orf283 (probable)	84	70

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					[Bacillus ubtilis]		
121	2	1274	2119	gi 290643	ATPase [Enterococcus hirae]	84	67
121	6	5016	5219	gi 153765	DNA polymerase I [Streptococcus pneumoniae]	84	66
128	27	22456	20453	gi 437916	isoleucyl-tRNA synthetase [Staphylococcus aureus]	84	71
130	1	2	133	gi 1237013	ORF2 [Bacillus subtilis]	84	74
138	35	26712	25777	gi 143795	transfer RNA-Tyr synthetase [Bacillus subtilis]	84	69
164	28	26378	27277	gnl PID e247026	orf6 [Lactobacillus sake]	84	72
171	1	158	2719	gi 499335	secA protein [Staphylococcus carnosus]	84	68
210	5	4870	3884	gi 950062	hypothetical yeast protein 1 [Mycoplasma capricolum] pir S48578 S48578 hypothetical protein - Mycoplasma capricolum SGC3 (fragment)	84	75
217	7	5222	3546	gi 143597	CTP synthetase [Bacillus subtilis]	84	68
243	1	1088	126	gi 143608	sporulation protein [Bacillus subtilis]	84	70
275	1	578	48	gi 1103865	formyl-tetrahydrofolate synthetase [Streptococcus mutans]	84	72
281	1	333	698	gi 1303962	YqjK [Bacillus subtilis]	84	68
292	23	18340	18038	gi 142988	membrane transport protein [Bacillus stearothermophilus] pir A42478 A42478 glutamine transport protein glnQ - Bacillus stearothermophilus	84	61
309	2	1114	722	gi 1644219	RNA polymerase beta' subunit [Bacillus subtilis]	84	72
315	1	668	3	gi 149601	thymidylate synthase (EC 2.1.1.45) [Lactobacillus casei]	84	72
334	6	5375	6862	gi 1354211	PEP12-like protein [Bacillus subtilis]	84	71
338	10	7585	10479	gi 467444	transcription-repair coupling factor [Bacillus subtilis] sp P37474 MFD_BACSU	84	68
					TRANSCRIPTION-REPAIR COUPLING FACTOR		

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
					(TRCF).		
338	14	12713	13018	gi 467448	unknown [Bacillus subtilis]	84	64
340	3	1068	2273	gi 40046	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] ir S15936 NUSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	84	69
375	2	1430	1780	gi 1402531	ORF10 [Enterococcus faecalis]	84	64
381	1	2	1279	gnl PID e208212	DNA topoisomerase IV [Streptococcus pneumoniae]	84	67
421	1	5	151	gi 710632	beta-glucosidase [Bacillus subtilis]	84	73
421	3	1229	1465	gi 710632	beta-glucosidase [Bacillus subtilis]	84	65
445	1	1080	190	gi 46985	glucose-1-phosphate thymidyltransferase [Salmonella enterica] ir S23342 S23342 hypothetical protein 6.1 - Salmonella choleraesuis p P55254 RFBA_SALAN GLUCOSE-1-PHOSPHATE THYMIDYLTRANSFERASE (EC 7.7.24) (DTDP-GLUCOSE SYNTHASE) (DTDP-GLUCOSE PYROPHOSPHO	84	71
466	9	10467	11006	gi 147403	mannose permease subunit II-P-Man [Escherichia coli]	84	61
497	2	469	1680	gi 1220529	methyl transferase [Streptococcus pneumoniae]	84	72
545	2	309	2171	gi 532548	ORF15 [Enterococcus faecalis]	84	68
550	5	2744	2265	gi 455528	ORF2 [Streptococcus thermophilus bacteriophage]	84	54
637	5	2679	3545	gnl PID e236571	cell wall anchoring signal [Enterococcus faecalis]	84	72
653	3	1023	736	gi 1408584	ltrC [Lactococcus lactis lactis]	84	72
674	1	763	254	gi 467452	unknown [Bacillus subtilis]	84	66
788	1	165	500	gi 1196907	daunorubicin resistance protein [Streptomyces peuceetius]	84	66

Table 3: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
796	1	1	366	gi 496504	orf beta [Streptococcus pyogenes]	84	67
812	1	2	415	gi 511075	ORF2 [Streptococcus agalactiae]	84	73
935	2	1317	949	gnl PIDe247026	orf6 [Lactobacillus sake]	84	73
954	1	3	470	gi 40019	ORF 821 (aa 1-821) [Bacillus subtilis]	84	67
17	3	2922	3311	gi 532555	ORF22 [Enterococcus faecalis]	83	69
17	12	8919	10130	gi 532553	ORF20 [Enterococcus faecalis]	83	64
17	30	30339	29137	gi 467416	unknown [Bacillus subtilis]	83	69
22	4	3208	3453	gi 467469	unknown [Bacillus subtilis]	83	64
28	3	6158	3471	pir A26738 SYBSVS	valine--tRNA ligase (EC 6.1.1.9) - Bacillus stearothermophilus	83	70
75	2	359	1405	gi 310628	phosphoenolpyruvate:sugar phosphotransferase system enzyme I Streptococcus mutans]	83	72
78	4	6971	5841	gi 155571	alcohol dehydrogenase I (adhA) (EC 1.1.1.1) [Zymomonas mobilis]	83	72
95	9	6859	7521	gi 487280	pir A35260 A35260 alcohol dehydrogenase (EC 1.1.1.1) I - Zymomonas obilis	83	66
98	3	2785	4008	gi 984803	Na+ -ATPase subunit D [Enterococcus hirae]	83	71
107	3	1467	988	sp P37214 ERA_STRM U	ATPase [Bacillus subtilis]	83	73
122	4	2781	3047	gi 467436	GTP-BINDING PROTEIN ERA HOMOLOG.	83	60
128	3	1572	2633	gi 559471	unknown [Bacillus subtilis]	83	64
128	34	28154	26844	gi 142941	pyruvate, orthophosphate dikinase [Mesembryanthemum crystallinum]	83	69
141	2	555	809	pir S03556 R3BS18	pir S49497 S49497 pyruvate, orthophosphate dikinase (EC 2.7.9.1) - ommon ice plant	83	71
173	8	6237	7241	gi 451216	ftsZ [Bacillus subtilis]	83	70

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
190	6	3124	2738	gi 532555	ORF22 [Enterococcus faecalis]	83	69
273	1	29	436	gnl PID e269878	ribosomal protein S15 [Bacillus subtilis]	83	71
334	1	3	920	gnl PID e248484	X-1 [Homo sapiens]	83	71
350	4	2723	2941	gi 511015	dihydroorotate dehydrogenase A [Lactococcus lactis] sp P54321 PYDA_LACLC DIHYDROOROTATE DEHYDROGENASE A (EC 1.3.3.1) DIHYDROOROTATE OXIDASE A) (DHODEHASE A).	83	66
367	2	1218	529	gi 142463	RNA polymerase alpha-core-subunit [Bacillus subtilis]	83	70
399	23	21538	22989	gi 40025	homologous to E.coli 50K [Bacillus subtilis]	83	67
399	25	25046	25879	gi 43939	D-glucitol-6-P-Dehydrogenase [Klebsiella pneumoniae] ir S50186 S50186 sorbitol-6-phosphate 2-dehydrogenase (EC 1.1.1.140) - Klebsiella pneumoniae	83	61
401	7	5097	5864	gi 755153	ATP-binding protein [Bacillus subtilis]	83	64
438	2	217	681	gi 530798	LysB [Bacteriophage phi-LC3]	83	67
497	4	2191	3402	gi 532553	ORF20 [Enterococcus faecalis]	83	63
539	2	2182	782	gi 467475	unknown [Bacillus subtilis]	83	67
563	1	2	1084	gi 142521	deoxyribodipyrimidine photolyase [Bacillus subtilis] gnl PID e255102 deoxyribodipyrimidine photolyase [Bacillus subtilis]	83	68
565	4	1018	1206	gi 1123066	weak similarity to bovine cAMP-dependant protein kinase II-B-binding protein (PIR:A39782) [Caenorhabditis elegans]	83	58
577	1	1	561	gi 1303854	YggG [Bacillus subtilis]	83	63
635	3	1210	1527	gi 1402526	ORF5 [Enterococcus faecalis]	83	65
644	1	2	442	gi 153801	enzyme scr-II [Streptococcus mutans]	83	69
655	3	848	1246	gi 147404	mannose permease subunit II-M-Man [Escherichia coli]	83	66

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
675	1	1	621	gi 467470	lysY-trNA thynthetase [Bacillus subtilis]	83	71
763	2	374	640	gi 145851	envM [Escherichia coli]	83	61
774	1	658	2	gi 1256145	ybbP [Bacillus subtilis]	83	60
3	1	58	327	gi 312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	82	70
5	10	6389	7708	sp P30053 SYH_STRE Q	HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) (HISRS)	82	71
27	4	1906	1145	gi 1303960	YqjI [Bacillus subtilis]	82	71
32	2	1333	965	gi 1303839	YqfR [Bacillus subtilis]	82	60
34	1	1643	324	gnl PID e218042	pyruvate kinase [Lactobacillus delbrueckii]	82	68
55	9	4182	5054	gi 1685110	tetrahydrofolate dehydrogenase/cyclohydrolase [Streptococcus thermophilus]	82	70
62	7	4644	4210	gi 143723	putative [Bacillus subtilis]	82	66
88	2	995	1624	gi 535349	CodW [Bacillus subtilis]	82	66
94	7	4790	3432	gi 1146247	asparaginyl-tRNA synthetase [Bacillus subtilis]	82	67
110	23	21590	20742	gi 467403	serYl-trNA synthetase [Bacillus subtilis]	82	69
114	7	8623	9228	gi 703442	thymidine kinase [Streptococcus gordonii]	82	68
123	6	4499	4996	gi 467356	unknown [Bacillus subtilis]	82	68
130	3	1413	2381	gi 308851	ATP binding protein [Lactococcus lactis]	82	64
144	3	3292	2339	gnl PID e183449	putative ATP-binding protein of ABC-type [Bacillus subtilis]	82	62
144	7	5331	5110	gi 335495	A23R; putative [Vaccinia virus]	82	47
159	4	2533	5010	gi 143148	transfer RNA-Leu synthetase [Bacillus subtilis]	82	71
159	6	5845	5387	gi 467354	unknown [Bacillus subtilis]	82	55
171	8	8510	9349	gi 1591672	phosphate transport system ATP-binding protein [Methanococcus jannaschii]	82	61
222	5	2158	3402	gi 143444	RNase PH [Bacillus subtilis]	82	66



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
254	6	1621	1112	gi 49316	ORF2 gene product [Bacillus subtilis]	82	61
279	12	9839	8442	gi 1237019	Srb [Bacillus subtilis]	82	67
288	1	22	546	gi 149393	IecA [Lactococcus lactis]	82	73
345	8	5608	8118	gi 442360	ClpC adenosine triphosphatase [Bacillus subtilis]	82	63
367	3	1472	1110	gi 142463	RNA polymerase alpha-core-subunit [Bacillus subtilis]	82	75
367	9	4961	3660	gi 44073	SecY protein [Lactococcus lactis]	82	65
367	28	12719	12411	pir A02815 R5B523	ribosomal protein L23 - Bacillus stearothermophilus	82	66
367	29	13330	12701	gi 1165304	L4 [Bacillus subtilis]	82	67
379	5	4396	3107	gi 887820	UUG start; possible frameshift at end? [Escherichia coli]	82	71
393	2	1145	711	gi 1303993	YqkL [Bacillus subtilis]	82	67
416	1	3	650	gi 475113	sucrase [Pediococcus pentosaceus]	82	69
477	1	1	1209	gi 309663	signaling protein [Plasmid pCF10]	82	62
497	7	3760	4275	gi 532551	ORF18 [Enterococcus faecalis]	82	67
535	3	4275	1666	gi 1747434	KdpD [Clostridium acetobutylicum]	82	62
587	1	488	108	gi 1303840	YqfS [Bacillus subtilis]	82	71
623	2	122	1348	gi 460259	enolase [Bacillus subtilis]	82	67
656	1	1	1908	gi 1184680	polynucleotide phosphorylase [Bacillus subtilis]	82	69
687	1	227	1252	gi 40218	PRPP synthetase (AA 1-317) [Bacillus subtilis]	82	64
728	1	3	527	gi 1146183	putative [Bacillus subtilis]	82	65
741	1	3	704	gi 153804	sucrose-6-phosphate hydrolase [Streptococcus mutans]	82	66
846	1	458	3	gnl PID e221400	tex gene product [Bordetella pertussis]	82	76
865	1	18	308	gi 416006	orf CJ01.2 [Campylobacter jejuni]	82	57
876	1	207	689	gi 1064795	function unknown [Bacillus subtilis]	82	62
925	1	436	128	gi 1773195	hypothetical [Escherichia coli]	82	74

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
983	2	280	474	gi 40026	homologous to E.coli gida [Bacillus subtilis]	82	78
12	3	4778	5788	gi 1100074	tryptophanyl-tRNA synthetase [Clostridium longisporum]	81	68
31	4	2984	4456	gi 849026	hypothetical 54.6-kDa protein [Bacillus subtilis]	81	68
34	6	6707	6910	gi 606067	ORF_f444 [Escherichia coli]	81	54
37	1	1	144	gi 1303854	YqgG [Bacillus subtilis]	81	59
37	3	2671	1958	gi 40056	phoP gene product [Bacillus subtilis]	81	61
57	3	1733	3220	gi 1657506	hypothetical protein [Escherichia coli]	81	66
60	5	5564	4440	gi 143370	phosphoribosylpyrophosphate amidotransferase (PUR-F; EC 2.4.2.14) [Bacillus subtilis]	81	63
73	3	2706	1450	gi 853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]	81	61
88	4	1977	2732	gnl PID e137596	heat shock induced protein HtpO [Lactobacillus leichmannii]	81	67
88	5	2723	3040	gi 535350	CodX [Bacillus subtilis]	81	65
101	4	3091	2435	gi 1109687	ProZ [Bacillus subtilis]	81	60
101	7	5884	4661	gi 1109684	ProV [Bacillus subtilis]	81	64
101	9	7501	7965	gi 1001768	queuosine biosynthesis protein QueA [Synecocystis sp.]	81	47
116	5	2766	3395	gi 1146234	dihydrodipicolinate reductase [Bacillus subtilis]	81	66
121	5	4811	5074	gi 153765	DNA polymerase I [Streptococcus pneumoniae]	81	64
121	7	5203	7488	gi 153765	DNA polymerase I [Streptococcus pneumoniae]	81	70
127	5	5103	3826	gi 290561	ol188 [Escherichia coli]	81	48
147	1	299	1279	gi 467462	cysteine synthetase A [Bacillus subtilis]	81	65
147	2	1370	1861	gnl PID e281583	hypothetical 16.4 kd protein [Bacillus subtilis]	81	63

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
154	1	168	638	gi 149533	subtilis]		
154	2	1074	1277	gnl PID e242898	conjugated bile acid hydrolase [Lactobacillus plantarum]	81	66
158	14	13790	12324	gi 558559	aBIR [Lactococcus lactis]	81	59
164	5	2469	3035	gi 727436	pyrimidine nucleoside phosphorylase [Bacillus subtilis]	81	71
223	8	5293	6153	gnl PID e254976	putative 20-kDa protein [Lactococcus lactis]	81	61
238	1	185	937	gi 622991	hypothetical protein [Bacillus subtilis]	81	66
276	7	3109	2819	pir A41207 A41207	mannitol transport protein [Bacillus stearothermophilus] sp P50852 PTMB_BACST PTS SYSTEM, MANNITOL-SPECIFIC IIBC COMPONENT EIIBC-MTL) (MANNITOL- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE NYZME II, BC COMPONENT) (EC 2.7.1.69) (EII-MTL).	81	68
307	2	1983	3617	gi 153742	collagen 13, nonfibrillar - freshwater sponge [Ephydatia muelleri] (fragment)	81	77
322	2	122	286	gi 296147	dextran glucosidase [Streptococcus mutans]	81	69
326	6	5352	4513	gi 40041	Asd protein [Bacillus subtilis]	81	63
329	3	1774	1448	gi 1117994	pyruvate dehydrogenase (lipoamide) [Bacillus stearothermophilus]	81	69
346	3	1056	1199	gi 536970	ir S10798 DEBSPF pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) pha chain - Bacillus stearothermophilus		
362	4	1131	2213	gi 1001826	surface antigen A variant precursor [Streptococcus pneumoniae]	81	72
391	3	1345	575	gi 1184967	ORF_f543 [Escherichia coli]	81	43
441	3	1873	3447	gi 1742675	cadmium-transporting ATPase [Synechocystis sp.]	81	64
					ScrR [Streptococcus mutans]	81	66
					Phosphotransferase system enzyme II (EC 2.7.1.69) MalX [Escherichia coli]	81	64

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
556	2	1062	493	gi 1553037	RecN [Bacillus subtilis]	81	66
710	2	361	816	gi 1303840	YqfS [Bacillus subtilis]	81	68
804	1	403	2	gi 149533	conjugated bile acid hydrolase [Lactobacillus plantarum]	81	68
5	7	3311	4255	gi 407881	stringent response-like protein [Streptococcus equisimilis]	80	62
				pir S39975 S39975 stringent response-like protein - Streptococcus quisiimilis			
17	10	8283	8438	gi 1326394	B0218.7 gene product [Caenorhabditis elegans]	80	53
17	15	12258	12776	gi 532551	ORF18 [Enterococcus faecalis]	80	63
22	1	3	2180	gi 44027	Tma protein [Lactococcus lactis]	80	70
37	6	3707	5140	pir B47154 B47154	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	80	64
42	1	2	259	gi 1066157	chaperonin-10 [Thermus aquaticus thermophilus]	80	66
49	16	11106	11309	gi 1136430	similar to hypothetical protein YM9959.11C of S.cerevisiae. [Homo sapiens]	80	53
60	4	4465	3407	gi 143371	phosphoribosyl aminoimidazole synthetase (PUR-M) [Bacillus subtilis]	80	62
				pir H29326 AJBSCL			
				phosphoribosylformylglycinamide cyclo-ligase EC 6.3.3.1) - Bacillus subtilis			
60	9	9023	8745	pir E29326 E29326	hypothetical protein (pur operon) - Bacillus subtilis	80	50
66	1	1	783	gi 520753	DNA topoisomerase I [Bacillus subtilis]	80	66
80	3	2519	1821	gnl PID e236074	beta-phosphoglucomutase [Lactococcus lactis]	80	62
83	9	6268	5378	gi 1070079	R08B4.1 [Caenorhabditis elegans]	80	72
89	18	19093	18845	gi 39451	type III restriction endonuclease [Bacillus cereus] ir S15518 JC1116 type III site-specific deoxyribonuclease (EC	80	72

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
97	1	366	4	gi 148506	1.21.5) - Bacillus cereus (fragment)		
107	2	1094	591	sp P37214 ERA_STRM U	dnaJ [Erysipelothrix rhusiopathiae] GTP-BINDING PROTEIN ERA HOMOLOG.	80	70
114	3	1474	5076	gi 43863	pyruvate-flavodoxin oxidoreductase [Klebsiella pneumoniae] ir S01997 QKXBP pyruvate (flavodoxin) dehydrogenase (EC 1.2.99.-) Klebsiella pneumoniae	80	64
117	3	1456	2367	gi 40031	spoJ93 gene product [Bacillus subtilis]	80	62
126	3	1857	709	gi 551854	ORF2 [Erwinia herbicola]	80	56
128	28	23265	22447	gi 437916	isoleucyl-tRNA synthetase [Staphylococcus aureus]	80	68
133	10	9128	9856	gi 520844	orf4 [Bacillus subtilis]	80	63
158	4	3926	2703	gi 944943	phosphopentomutase [Bacillus subtilis]	80	64
172	5	3732	3920	sp P20182 YT14_STR FR	HYPOTHETICAL 29.1 KD PROTEIN IN TRANSPONSON TN4556.	80	63
180	16	15548	16393	gi 1773200	hypothetical protein [Escherichia coli]	80	66
181	10	8597	7407	gi 143806	AroF [Bacillus subtilis]	80	64
194	4	1580	1957	gi 47394	5-oxoprol-1-peptidase [Streptococcus pyogenes]	80	66
213	5	3515	4078	gnl PID e199384	pyrR gene product [Lactobacillus plantarum]	80	65
217	11	7724	8395	gi 1561567	Unknown [Bacillus subtilis]	80	65
218	6	4843	5331	gi 1574120	branched-chain-amino-acid transaminase [Haemophilus influenzae]	80	64
225	8	6092	5829	gi 530459	similar to phosphotransferase EII [Mycoplasma capricolum]	80	52
229	2	1170	178	gi 1502419	PisX [Bacillus subtilis]	80	59
243	3	2545	2150	gi 1732315	transport system permease homolog [Listeria monocytogenes]	80	64
275	2	694	939	gi 1256629	cold-shock protein [Bacillus subtilis]	80	65

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
307	3	3607	3888	gi 1321625	exo-alpha-1, 4-glucosidase [Bacillus stearothermophilus]	80	73
322	3	284	1090	gi 142828	aspartate semialdehyde dehydrogenase [Bacillus subtilis] sp Q04797 DHAS_BACSU ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (EC .2.1.11) (ASA DEHYDROGENASE).	80	62
349	1	2	616	gi 495089	recombinase [Staphylococcus aureus]	80	65
367	7	3511	2924	gi 44074	adenylate kinase [Lactococcus lactis]	80	64
386	7	4305	5306	gi 149396	lacD [Lactococcus lactis]	80	64
394	3	2642	3757	pir B39096 B39096	alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis	80	64
399	17	12070	13488	gi 1591862	oxaloacetate decarboxylase, alpha subunit [Methanococcus jannaschii]	80	61
399	24	22979	24907	gi 40026	homologous to E.coli gidA [Bacillus subtilis]	80	67
435	3	2217	2032	gi 559863	clyA [Plasmid pAD1]	80	78
466	1	3	1208	gi 467330	replicative DNA helicase [Bacillus subtilis]	80	61
475	4	3402	2947	gi 532547	ORF14 [Enterococcus faecalis]	80	68
491	4	3844	4392	gi 473892	large-conductance mechanosensitive channel [Escherichia coli] gi 473420 yhcC [Escherichia coli]	80	56
605	2	1252	338	gi 580875	ipa-57d gene product [Bacillus subtilis]	80	69
615	1	760	14	gi 467469	unknown [Bacillus subtilis]	80	66
668	1	117	587	pir S16974 R5BS7F	ribosomal protein L9 - Bacillus stearothermophilus	80	71
684	2	694	464	gi 786314	Highly similar to Glycogen debranching enzyme 4-alpha-glucanotransferase, Swiss Prot. accession number P35573) Saccharomyces cerevisiae]	80	33
767	1	1	480	gi 41828	istB gene product [Escherichia coli]	80	52
818	1	1	357	gi 1743856	intrageneric coaggregation-relevant	80	66

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
833	1	325	95	gi 1561567	adhesin [Streptococcus gordonii]		
934	1	394	56	gi 1001706	Unknown [Bacillus subtilis]	80	68
948	1	465	4	gi 1773196	ABC transporter subunit [Synechocystis sp.]	80	63
949	1	61	411	gi 1330380	similar to B. stearothermophilus N-carbamyl-L-amino acid amidohydrolase [Escherichia coli]	80	59
20	2	468	1262	gi 1256698	Similar to cystathionine gamma-lyase [Caenorhabditis elegans]	80	61
22	3	2420	3238	gi 467460	chitinase [Serratia marcescens]	79	67
24	1	39	1109	gi 1303821	unknown [Bacillus subtilis]	79	59
26	1	214	873	gi 403984	YgfE [Bacillus subtilis]	79	61
47	8	10268	8106	gi 153657	deoxyguanosine kinase/deoxyadenosine kinase(I) subunit Lactobacillus acidophilus	79	68
48	9	9905	9198	gi 290566	mismatch repair protein [Streptococcus pneumoniae] pir A33589 A33589 mismatch repair protein hexB - Streptococcus pneumoniae	79	63
58	4	4677	3694	gi 1653179	f213 [Escherichia coli]	79	53
63	6	3605	5443	gi 1064813	hydrogenase subunit [Synechocystis sp.]	79	52
88	8	5493	4771	gnl PID e208252	homologous to sp:PHOR_BACSU [Bacillus subtilis]	79	55
146	8	6649	5609	gi 153676	unidentified [Streptococcus pneumoniae]	79	57
149	4	2554	1976	gi 1216490	tagatose 1,6-aldolase [Streptococcus mutans]	79	63
158	2	1859	1143	gi 1276873	DNA/pantothenate metabolism flavoprotein [Streptococcus mutans]	79	64
179	19	19022	18417	gi 467372	DeoD [Streptococcus thermophilus]	79	67
222	2	982	230	gi 142988	3'-exo-deoxyribonuclease [Bacillus subtilis]	79	61
					membrane transport protein [Bacillus subtilis]	79	59

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					stearothermophilus] pir A42478 A42478 glutamine transport protein glnQ - Bacillus tearothermophilus		
228	6	4060	3401	gi 413950	ipa-26d gene product [Bacillus subtilis]	79	55
229	3	3270	1219	gnl PID el86699	MnsA [Streptococcus pneumoniae]	79	62
238	7	5750	5100	gi 596046	L8003.16 gene product [Saccharomyces cerevisiae]	79	55
269	10	6664	5489	gi 1303788	YgeH [Bacillus subtilis]	79	63
274	1	1	1143	gi 153062	helicase [Staphylococcus aureus]	79	65
290	9	7364	8779	gi 466882	psl1; B1496_C2_189 [Mycobacterium leprae]	79	64
292	22	18122	17595	gi 1303951	YqiZ [Bacillus subtilis]	79	61
316	3	864	2003	gi 1146207	putative [Bacillus subtilis]	79	58
326	2	1772	360	gi 40044	dihydrolipoamide dehydrogenase [Bacillus stearothermophilus] ir S13839 S13839 dihydrolipoamide dehydrogenase (EC 1.8.1.4) - cillus stearothermophilus	79	65
363	5	5738	7180	gi 1657519	hypothetical protein [Escherichia coli]	79	63
367	11	5668	5447	gi 216337	ORF for L30 ribosomal protein [Bacillus subtilis]	79	63
375	5	4346	3393	gi 1644203	unknown [Bacillus subtilis]	79	62
406	2	666	1481	gi 49316	ORF2 gene product [Bacillus subtilis]	79	58
460	7	4973	5860	gi 1276664	acetyl-CoA carboxylase carboxytransferase beta subunit [Porphyra purpurea]	79	62
486	1	380	3	gi 1256618	transport protein [Bacillus subtilis]	79	63
488	3	987	1997	gi 532547	ORF14 [Enterococcus faecalis]	79	69
500	2	1358	681	gi 535662	transposase [Insertion sequence IS1251]	79	75
523	3	1803	820	gi 142981	ORF5; This ORF includes a region (aa23-103) containing a potential non-sulphur centre homologous to a region of Rhodospirillum rubrum and Chromatium vinosum; putative [Bacillus stearothermophilus] pir PQ0299 PQ0299	79	62



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					hypothetical protein 5 (gldA 3' region) -		
552	2	2401	902	gi 887851	ORF_0479 [Escherichia coli]	79	63
587	2	622	434	gi 1303840	Yqfs [Bacillus subtilis]	79	66
612	1	1	378	gi 1064791	function unknown [Bacillus subtilis]	79	56
654	1	2	286	pir A47079 A47079	heat shock protein DnaJ - Lactococcus lactis	79	75
701	2	325	534	gi 143793	tyrosyl-tRNA synthetase [Bacillus caldovenax]	79	63
708	2	369	566	gi 488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	79	66
840	1	140	1078	gi 1573250	aspartate aminotransferase (aspC) [Haemophilus influenzae]	79	65
5	9	5555	6049	gi 407880	ORF1 [Streptococcus equisimilis]	78	58
33	4	3755	4597	gi 1742846	NH(3)-dependent NAD(+) synthetase (EC 6.3.5.1) (Nitrogen-regulatory protein). [Escherichia coli]	78	64
60	7	8100	5854	gi 143369	phosphoribosylformyl glycine synthetase II (PUR-Q) [Bacillus subtilis]	78	62
65	4	3407	2625	gi 1661179	high affinity branched chain amino acid transport protein [Streptococcus mutans]	78	67
76	7	5760	4747	gi 1161061	dioxygenase [Methylobacterium extorquens]	78	62
81	11	7141	6824	gi 1072380	ORF3 [Lactococcus lactis]	78	67
83	5	2559	2843	gi 1256896	L9606.1 gene product [Saccharomyces cerevisiae]	78	52
85	4	4298	3288	gi 142612	branched chain alpha-keto acid dehydrogenase E1-beta [Bacillus subtilis]	78	61
85	8	6723	6307	gi 1303941	YgiV [Bacillus subtilis]	78	62
88	10	6477	6689	gi 222595	nucleocapsid protein [Sialodacryoadenitis virus]	78	57
93	5	1838	2641	gi 405133	putative [Bacillus subtilis]	78	51
117	1	3	707	gi 40027	homologous to E.coli gidB [Bacillus]	78	64

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
117	11	9624	8338	gi 467403	subtilis]		
132	2	2323	2024	gi 683484	seryl-tRNA synthetase [Bacillus subtilis]	78	63
133	3	2241	3413	gi 405622	fusion protein [Mumps virus]	78	63
150	2	568	1425	gnl PID e185373	unknown [Bacillus subtilis]	78	63
155	2	604	1182	gi 285628	ceuD gene product [Campylobacter coli]	78	52
					transcription antitermination factor NusG [Bacillus subtilis] pir S39859 S39859	78	61
					transcription antitermination factor NusG - acillus subtilis		
156	2	308	2629	gi 1573874	ATP-dependent protease binding subunit (clpB) [Haemophilus influenzae]	78	59
158	3	2719	1868	gi 1638804	purine nucleoside phosphorylase [Bacillus stearothermophilus]	78	64
160	5	2058	3050	gi 1161061	dioxygenase [Methylobacterium extorquens]	78	60
161	3	1466	3295	gnl PID e280490	unknown [Streptococcus pneumoniae]	78	62
169	1	2	2206	gi 1072361	pyruvate-formate-lyase [Clostridium pasteurianum]	78	61
171	2	2833	3897	sp P28367 RF2_BACS U	PROBABLE PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2) (FRAGMENT).	78	64
180	15	14851	15567	gi 1773199	hypothetical protein [Escherichia coli]	78	67
185	1	1142	3	pir C33496 C33496	hisc homolog - Bacillus subtilis	78	59
188	3	1863	4178	gnl PID e256969	nifJ gene product [Enterobacter agglomerans]	78	62
216	7	5136	5600	gnl PID e276830	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]	78	60
216	8	5531	6508	gnl PID e276830	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]	78	63
238	26	24515	25387	gi 396681	rhamnose-1-phosphate aldolase [Escherichia coli]	78	56
256	6	4189	6237	gi 467427	methionyl-tRNA synthetase [Bacillus subtilis]	78	67

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
292	4	2063	2353	gi 1742823	subtilis] Proton/sodium-glutamate symport protein (Glutamate-aspartate carrier protein). [Escherichia coli]	78	62
305	1	268	1872	gi 143582	spoIIIEA protein [Bacillus subtilis]	78	58
337	2	2332	1448	gi 308861	GTG start codon [Lactococcus lactis]	78	63
338	2	606	1466	gi 1773142	similar to the 20.2kd protein in TETB-EXOA region of B. subtilis [Escherichia coli]	78	66
362	1	109	429	gi 150719	cadmium resistance protein [Plasmid pI258]	78	51
379	3	2878	1922	gi 887824	ORF_o310 [Escherichia coli]	78	60
446	2	962	1636	gi 537235	Kenn Rudd identifies as gpMB [Escherichia coli]	78	43
495	5	3038	3502	gi 634107	kdpB [Escherichia coli]	78	58
502	3	3077	1470	gi 1652592	peptide-chain-release factor 3 [Synechocystis sp.]	78	58
523	1	2	616	gi 289288	lexA [Bacillus subtilis]	78	59
571	1	99	365	gnl PID e249644	YneF [Bacillus subtilis]	78	65
573	3	1258	1971	gi 1731683	component II of heptaprenyl diphosphate synthase [Bacillus stearothermophilus]	78	50
575	2	434	168	gi 58831	The experimental evidence that this sequence codes for a complete gag otein is that transfection of the viral genome results in oduction of infectious virus [Cas-Br-E murine leukemia virus] p P27460 GAG_MLVCB GAG POLYPROTEIN (CONTAINS: CORE PROTEIN P15; N	78	47
607	1	148	708	gi 530410	Ala-tRNA synthetase [Mycoplasma capricolum]	78	63
655	2	300	899	gi 147404	mannose permease subunit II-M-Man [Escherichia coli]	78	60
704	1	181	2	gi 467430	unknown [Bacillus subtilis]	78	63
708	1	1	378	gi 443985	alcohol dehydrogenase [Entamoeba	78	61

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					histolytica]		
732	1	661	2	gi 1064791	function unknown [Bacillus subtilis]	78	55
785	1	2	679	gi 556014	UDP-N-acetyl muramate-alanine ligase [Bacillus subtilis]	78	59
786	1	2	172	gi 536992	SugES [Escherichia coli]	78	60
820	2	1602	1144	gi 153749	UDPglucose 4-epimerase [Streptococcus thermophilus] pir A44509 A44509 UDPglucose 4-epimerase (EC 5.1.3.2) - treptococcus thermophilus	78	60
887	1	337	2	gi 495046	tripeptidase [Lactococcus lactis]	78	70
970	2	395	234	gi 1652190	Fat protein [Synechocystis sp.]	78	51
4	7	6069	5656	gi 1573482	high affinity ribose transport protein (rbsD) [Haemophilus influenzae]	77	51
45	16	12065	14047	gi 666069	orf2 gene product [lactobacillus leichmannii]	77	51
49	13	8199	9992	gnl PID e228615	homologous to yqcC of the skin element [Bacillus subtilis]	77	59
60	2	2895	1300	gi 143373	phosphoribosyl aminoimidazole carboxy formyl ormyltransferase/inosine monophosphate cyclohydrolase (PUR-H(J)) Bacillus subtilis]	77	63
70	6	5118	3874	gi 912464	No definition line found [Escherichia coli]	77	53
70	7	5172	5756	gi 288413	glutamate dehydrogenase (NADP+) [Corynebacterium glutamicum] pir S32227 S32227 glutamate dehydrogenase (NADP+) (EC 1.4.1.4) - orynebacterium glutamicum	77	65
74	10	7303	5864	gi 289284	cysteinyI-tRNA synthetase [Bacillus subtilis]	77	62
74	12	9559	8078	gi 289282	glutamyl-tRNA synthetase [Bacillus subtilis]	77	57

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
88	6	3013	3843	gi 535351	CodY [Bacillus subtilis]	77	57
89	6	5749	2510	gi 1695686	pyruvate carboxylase [Bacillus stearothermophilus]	77	62
91	1	396	728	gi 1184044	L-glutamine:D-fructose-6-P amidotransferase precursor [Thermus aquaticus thermophilus]	77	66
98	4	3992	5710	gi 984804	transmembrane protein [Bacillus subtilis]	77	56
124	1	2	940	gnl PID e199002	prolidase PepQ [Lactobacillus delbrueckii]	77	60
158	5	4845	4171	gi 435297	unknown [Lactococcus lactis]	77	48
162	6	7426	5882	gi 142992	glycerol kinase (glpK) (EC 2.7.1.30) [Bacillus subtilis] pir B45868 B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp P18157 GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL - PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK).	77	60
164	1	179	1102	gi 882532	ORF_o294 [Escherichia coli]	77	57
164	22	24158	23646	gi 1573564	hypothetical [Haemophilus influenzae]	77	36
171	6	6656	7639	gi 1303855	YqgH [Bacillus subtilis]	77	59
171	9	9198	9683	gi 1591672	phosphate transport system ATP-binding protein [Methanococcus jannaschii]	77	57
202	4	2967	3422	gi 147782	ruvA protein (gtg start) [Escherichia coli]	77	50
202	6	3662	4693	gi 147783	ruvB protein [Escherichia coli]	77	58
213	1	3	1046	gi 1103865	formyl-tetrahydrofolate synthetase [Streptococcus mutans]	77	63
217	10	6870	7742	gi 414014	ipa-90d gene product [Bacillus subtilis]	77	50
223	5	4171	4902	gnl PID e254974	autolysin response regulator [Bacillus subtilis]	77	55
223	7	5024	5473	gnl PID e254975	hypothetical protein [Bacillus subtilis]	77	58
228	10	7747	6035	gi 467409	DNA polymerase III subunit [Bacillus subtilis]	77	61
229	15	16711	14261	gnl PID e290286	priA [Bacillus subtilis]	77	62

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
232	3	1742	1437	gi 142708	comG3 gene product [Bacillus subtilis]	77	50
238	25	23174	24511	pir B48649 B48649	L-rhamnose isomerase (EC 5.3.1.14) - Escherichia coli	77	59
238	32	29472	28708	gi 451072	di-tripeptide transporter [Lactococcus lactis]	77	56
244	4	3591	2809	gi 1773173	similar to M. jannaschii MJ0938 [Escherichia coli]	77	60
269	5	3890	3522	gi 1303793	YgeL [Bacillus subtilis]	77	55
276	6	2840	2328	pir PC1127 PC1127	hypothetical 110 protein (lytA 5' region) - Lactococcus lactis phage US3 (fragment)	77	50
291	1	119	916	gi 556014	UDP-N-acetyl muramate-alanine ligase [Bacillus subtilis]	77	63
304	2	941	2020	gnl PID e285001	CTORF239 [Staphylococcus aureus]	77	62
305	4	3618	4394	gi 709993	hypothetical protein [Bacillus subtilis]	77	54
327	8	5697	6005	gi 153570	H+ ATPase [Enterococcus faecalis]	77	61
341	4	1206	1937	gi 1303951	YqiZ [Bacillus subtilis]	77	62
360	1	429	4	gi 897754	nonstructural protein NSP3 [Human rotavirus]	77	38
362	3	541	1239	gi 1001826	cadmium-transporting ATPase [Synechocystis sp.]	77	60
363	9	13917	12652	gi 1574390	C4-dicarboxylate transport protein [Haemophilus influenzae]	77	55
367	14	7218	6679	pir A02766 R5BS0F	ribosomal protein L6 - Bacillus stearothermophilus	77	63
386	8	5456	5776	gnl PID e281578	hypothetical 12.2 kd protein [Bacillus subtilis]	77	61
394	4	3706	4167	pir B39096 B39096	alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis	77	55
402	1	710	3	gi 533105	unknown [Bacillus subtilis]	77	59
408	2	1357	584	gi 666983	putative ATP binding subunit [Bacillus subtilis]	77	58

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
460	6	3562	4938	gi 1055246	biotin carboxylase [Bacillus subtilis]	77	60
466	7	8657	9253	gi 147402	mannose permease subunit III-Man [Escherichia coli]	77	61
475	5	3794	3234	gi 532547	ORF14 [Enterococcus faecalis]	77	68
498	1	1	603	gi 410137	ORFX13 [Bacillus subtilis]	77	58
515	1	107	574	gi 1303815	YgeY [Bacillus subtilis]	77	60
518	6	2980	4518	gi 1402515	membrane-spanning transporter protein [Clostridium perfringens]	77	56
523	5	2527	2333	gi 149601	thymidylate synthase (EC 2.1.1.45) [Lactobacillus casei]	77	66
526	2	1782	436	gi 1750124	xylose isomerase [Bacillus subtilis]	77	62
552	7	6809	6135	gi 534045	antiterminator [Bacillus subtilis]	77	51
607	3	778	936	gi 1015321	alanyl-tRNA synthetase [Homo sapiens]	77	51
624	3	2289	2555	gnl PID e187971	orf121 gene product [Lactococcus lactis]	77	57
781	1	15	485	gi 580883	ipa-88d gene product [Bacillus subtilis]	77	65
850	2	895	572	gi 142520	thioredoxin [Bacillus subtilis]	77	59
853	1	186	4	gi 39962	ribosomal protein L35 (AA 1-66) [Bacillus stearothermophilus] ir S05347 R5BS35	77	66
					ribosomal protein L35 - Bacillus earothermophilus		
944	1	2	172	gi 425467	transposase [Lactobacillus helveticus]	77	50
10	1	1	258	gnl PID e234078	hom [Lactococcus lactis]	76	63
12	4	7650	5842	gnl PID e254877	unknown [Mycobacterium tuberculosis]	76	57
17	29	29022	28153	gi 1500003	mutator mutT protein [Methanococcus jannaschii]	76	47
23	15	8897	10285	gi 153960	ethanolamine ammonia-lyase (eutB) [Salmonella typhimurium] pir A36570 A36570	76	64
					ethanolamine ammonia-lyase (EC 4.3.1.7) 55K chain Salmonella typhimurium		
29	2	1024	500	gi 40011	ORF17 (AA 1-161) [Bacillus subtilis]	76	61
33	1	14	1552	gi 148304	beta-1,4-N-acetylmuramoylhydrolase	76	60

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					[Enterococcus hirae] pir A42296 A42296 lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus irae (ATCC 9790)		
34	7	7432	6965	gi 44067	ORF1 C-terminal [Lactococcus lactis]	76	59
45	8	3708	4166	gi 1303698	BltD [Bacillus subtilis]	76	56
47	9	12849	10270	gi 1002520	MutS [Bacillus subtilis]	76	59
55	8	3614	4105	gi 1303915	YqhZ [Bacillus subtilis]	76	53
55	11	6385	6642	gi 216583	ORF1 [Escherichia coli]	76	45
57	14	17283	16597	gi 1183887	integral membrane protein [Bacillus subtilis]	76	56
59	6	3112	2426	gi 392872	repressor protein [Pasteurella multocida]	76	47
64	1	1242	46	gi 483941	blt gene product [Bacillus subtilis]	76	55
67	3	1370	2146	gnl PID e199390	orotate phosphoribosyltransferase [Lactobacillus plantarum]	76	57
69	2	837	334	gi 1377831	unknown [Bacillus subtilis]	76	57
70	1	164	1588	gi 895751	putative 6-phospho-beta-glucosidase [Bacillus subtilis] pir S57762 S57762 probable 6-phospho-beta-glucosidase - Bacillus ubtilis	76	60
74	11	7826	7269	pir E53402 E53402	serine O-acetyltransferase (EC 2.3.1.30) - Bacillus stearothermophilus	76	54
74	13	10073	9588	gi 289281	unknown [Bacillus subtilis]	76	60
85	11	7809	7102	gi 457634	butyrate kinase [Clostridium acetobutylicum]	76	61
94	8	6036	4801	gi 142538	aspartate aminotransferase [Bacillus sp.]	76	57
94	14	17174	12801	gi 40060	DNA polymerase III (AA 1-1437) [Bacillus subtilis] p P13267 DP3A_BACSU DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7).	76	62
94	15	19140	17407	gi 1573733	prolyl-tRNA synthetase (proS) [Haemophilus influenzae]	76	54
95	1	1	1290	gi 472918	v-type Na-ATPase [Enterococcus hirae]	76	59



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
95	4	2367	3194	gi 487276	Na <sup>+</sup> -ATPase subunit C [Enterococcus hirae]	76	48
99	1	1	171	gi 1353874	unknown [Rhodobacter capsulatus]	76	52
100	5	5414	5064	gi 1591962	M. jannaschii predicted coding region MJ1322 [Methanococcus jannaschii]	76	46
100	27	23165	21198	gi 216151	DNA polymerase (gene L; ttg start codon) [Bacteriophage SPO2] gi 579197 SPO2 DNA polymerase (aa 1-648) [Bacteriophage SPO2] pir A21498 DJBPS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage PO2	76	62
106	1	1511	264	gi 1750108	Ynba [Bacillus subtilis]	76	61
116	4	2480	2854	gi 755602	unknown [Bacillus subtilis]	76	60
116	6	3299	3625	gi 1146234	dihydrodipicolinate reductase [Bacillus subtilis]	76	56
122	5	3029	3619	gi 467436	unknown [Bacillus subtilis]	76	52
123	10	9109	10389	gi 1773196	similar to B. stearothermophilus N-carbamyl-L-amino acid amidohydrolase [Escherichia coli]	76	61
124	5	4087	3182	gi 974332	NAD(P)H-dependent dihydroxyacetone-phosphate reductase [Bacillus ubtilis]	76	58
130	5	3341	4294	gi 308853	transmembrane protein [Lactococcus lactis]	76	55
132	3	2265	5117	gi 1673889	(AE000022) Mycoplasma pneumoniae, excinuclease ABC subunit A; similar to Swiss-Prot Accession Number P07671, from E. coli [Mycoplasma pneumoniae]	76	59
138	34	25849	25409	gi 143795	transfer RNA-Tyr synthetase [Bacillus subtilis]	76	56
139	1	3	350	gnl PID e191395	mobilisation protein [Lactococcus lactis]	76	65
141	1	2	544	gi 662792	single-stranded DNA binding protein [unidentified eubacterium]	76	64
155	9	7612	7058	gnl PID e247026	orf6 [Lactobacillus sake]	76	57
164	4	1889	2416	gi 727436	putative 20-kDa protein [Lactococcus lactis]	76	55

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
181	5	3475	2288	gi 1147744	PSR [Enterococcus hirae]	76	53
181	8	6281	4986	gi 683583	5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus lactis] pir S52580 S52580 3-phosphoshikimate 1-carboxyvinyltransferase (EC .5.1.19) - Lactococcus lactis	76	62
197	7	7662	8102	gi 1783253	homologous to many ATP-binding transport proteins; hypothetical [Bacillus subtilis]	76	58
222	16	10780	11298	gi 1591856	hypothetical protein (SP:P15889) [Methanococcus jannaschii]	76	64
229	1	1	138	gi 148316	NaH-antiporter protein [Enterococcus hirae]	76	47
233	6	3946	3341	gi 1591652	hypothetical protein (SP:P31065) [Methanococcus jannaschii]	76	60
238	2	844	1848	gi 622991	mannitol transport protein [Bacillus stearothermophilus] sp P50852 PTMB_BACST PTS SYSTEM, MANNITOL-SPECIFIC IIBC COMPONENT EIIBC-MTL) (MANNITOL- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE NYZME II, BC COMPONENT) (EC 2.7.1.69) (EII-MTL).	76	64
238	9	7235	7957	gi 1592142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	76	49
249	2	543	1235	gi 143156	membrane bound protein [Bacillus subtilis]	76	45
262	3	4131	2692	gnl PID e281591	catalase [Bacillus subtilis]	76	65
265	1	2	400	gi 141858	replication-associated protein [Plasmid pAD1]	76	52
271	13	8175	10844	gi 397973	Mg2+ transport ATPase [Salmonella typhimurium]	76	57
323	4	4128	4568	gnl PID e249023	T19B10.3 [Caenorhabditis elegans]	76	60
329	5	3270	2560	gi 310631	ATP binding protein [Streptococcus gordonii]	76	54
356	1	971	3	gi 971479	orf3 gene product [Lactobacillus]	76	52

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
					delbrueckii]		
371	1	1564	944	gi 1750125	xylose kinase [Bacillus subtilis]	76	57
375	6	5137	4238	gi 1644202	unknown [Bacillus subtilis]	76	58
382	2	508	2769	gi 442360	ClpC adenosine triphosphatase [Bacillus subtilis]	76	60
399	11	7811	8845	gi 1572970	acetate:SH-citrate lyase ligase (AMP) [Haemophilus influenzae]	76	54
399	13	9126	10034	gi 1572968	citrate lyase beta chain (acyl lyase subunit) (cite) [Haemophilus influenzae]	76	57
485	1	3	1262	gi 564018	dihydrofolate synthetase [Streptococcus pneumoniae]	76	54
486	2	970	344	gi 1256617	adenine phosphoribosyltransferase [Bacillus subtilis]	76	61
536	1	220	2	gi 437389	transposase [Lactococcus lactis]	76	59
552	3	3969	2491	gi 882609	6-phospho-beta-glucosidase [Escherichia coli]	76	63
634	2	697	918	gi 1022725	unknown [Staphylococcus haemolyticus]	76	52
684	3	1191	688	gi 1256653	DNA-binding protein [Bacillus subtilis]	76	65
752	1	1111	929	gi 407907	ORF2 [Staphylococcus xylosus]	76	46
822	1	548	237	gi 144313	6.0 kd ORF [Plasmid Cole1]	76	73
923	1	2	421	gi 153843	trypsin-resistant surface T6 protein (tee6) precursor [Streptococcus yogenes]	76	57
953	2	534	187	gi 1592339	hypothetical protein (PIR:S52522) [Methanococcus jannaschii]	76	44
965	2	564	343	gi 1098898	CTRP [Plasmodium falciparum]	76	69
7	4	3754	4161	gi 495046	tripeptidase [Lactococcus lactis]	75	61
25	1	2	580	gi 1575577	DNA-binding response regulator [Thermotoga maritima]	75	57
45	7	3090	3350	gi 1673663	(AE000003) Mycoplasma pneumoniae, E07_orf166 Protein [Mycoplasma pneumoniae]	75	35
47	6	7526	6957	gi 1673843	(AE000019) Mycoplasma pneumoniae, pilB	75	58

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					homolog; similar to GenBank Accession Number E64124, from H. influenzae [Mycoplasma pneumoniae]		
51	1	15	1520	sp P39168 ATMA_ECO LI	MG(2+) TRANSPORT ATPASE, P-TYPE 1 (EC 3.6.1.-)	75	58
54	11	3761	3579	gi 1504026	similar to C.elegans protein (Z37093) [Homo sapiens]	75	56
55	5	1648	2562	gi 1303901	YqHT [Bacillus subtilis]	75	58
56	8	5873	5358	gi 895749	putative cellobiose phosphotransferase enzyme II'' [Bacillus ubtilis]	75	49
58	2	2707	1916	gi 1658403	formate dehydrogenase alpha subunit [Moorella thermoacetica]	75	58
71	1	110	1429	gi 1304007	LysA [Bacillus subtilis]	75	58
74	5	3436	3074	gi 467433	unknown [Bacillus subtilis]	75	61
74	8	5491	4631	gi 467483	unknown [Bacillus subtilis]	75	60
77	1	3	992	gi 1653966	47 kD protein [Synechocystis sp.]	75	34
81	1	26	862	gi 1064809	homologous to sp:HTRA_ECOLI [Bacillus subtilis]	75	55
89	11	11651	9801	gi 1573881	hypothetical [Haemophilus influenzae]	75	51
96	3	2521	1643	gi 1531619	NodB [Rhizobium sp.]	75	54
98	9	11494	10199	gi 1573043	hypothetical [Haemophilus influenzae]	75	53
110	12	11326	10283	gi 1184121	auxin-induced protein [Vigna radiata]	75	51
117	13	11200	9944	gi 457635	vancomycin histidine protein kinase [Enterococcus faecium] gi 801884 vans [Transposon Tn1546]	75	51
122	6	3812	5206	gi 467439	temperature sensitive cell division [Bacillus subtilis]	75	59
128	12	8262	7921	gi 466473	cellobiose phosphotransferase enzyme II' [Bacillus tearothermophilus]	75	48
128	38	31848	30733	gi 216300	peptidoglycan synthesis enzyme [Bacillus subtilis] sp P37585 MURG_BACSU MURG PROTEIN UPD-N-ACETYLGLUCOSAMINE--N-	75	56

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					ACETYL-MURAMYL-PENTAPEPTIDE) PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE).		
129	2	1916	2134	gnl PID e267624	Unknown, highly similar to Pseudomonas putida 4-oxalocrotonate tautomerase [Bacillus subtilis]	75	47
130	4	2375	3343	gi 495179	transmembrane protein [Lactococcus lactis]	75	55
133	1	3	1514	gnl PID e254877	unknown [Mycobacterium tuberculosis]	75	54
158	13	12326	11634	gi 809660	deoxyribose-phosphate aldolase [Bacillus subtilis] pir S49455 S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis	75	66
162	13	14285	12543	gi 1653222	cation-transporting ATPase PacL [Synechocystis sp.]	75	60
170	2	1280	921	sp P07999 DHGB_BAC ME	GLUCOSE 1-DEHYDROGENASE B (EC 1.1.1.47).	75	62
171	7	7618	8523	gi 1303856	YqgI [Bacillus subtilis]	75	52
179	14	14668	15255	gi 457177	alkyl hydroperoxide reductase [Salmonella typhimurium] sp P19479 AHPC_SALTY ALKYL HYDROPEROXIDE REDUCTASE C22 PROTEIN (EC 1.11.1.17)	75	55
181	6	4470	3604	gi 683585	prephenate dehydratase [Lactococcus lactis]	75	49
191	1	183	560	gnl PID e261991	putative orf [Bacillus subtilis]	75	57
197	3	2117	3592	gi 1783250	homologous to cytochrome d ubiquinol oxidase subunit I; hypothetical [Bacillus subtilis]	75	60
215	3	2545	2201	gnl PID e284996	ORF136 [Staphylococcus aureus]	75	54
216	1	2	256	gi 153570	H+ ATPase [Enterococcus faecalis]	75	53
223	4	2406	4193	gi 862312	lytS gene product [Staphylococcus aureus]	75	56
227	5	3004	3567	gi 144729	butanol dehydrogenase [Clostridium acetobutylicum] sp Q04944 ADHA_CLOAB NADH-	75	53

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					DEPENDENT BUTANOL DEHYDROGENASE A (EC .1.1.-) (BDH I).		
228	9	6032	5700	gi 467410	unknown [Bacillus subtilis]	75	59
229	16	17081	16848	gi 207398	tropomyosin T class IVD alpha-3 [Rattus norvegicus]	75	42
238	8	6038	7237	gi 141927	cscB gene product [Alcaligenes eutrophus]	75	39
244	10	7795	7460	gi 467419	unknown [Bacillus subtilis]	75	56
247	1	7	1431	gi 577569	PepV [Lactobacillus delbrueckii]	75	54
250	5	3416	3201	gi 1580783	sperm receptor [Strongylocentrotus purpuratus]	75	50
256	1	2	562	gi 709991	hypothetical protein [Bacillus subtilis]	75	56
262	2	1031	2479	gi 142783	DNA photolyase [Bacillus firmus]	75	59
263	1	222	890	gi 148304	beta-1,4-N-acetylmuramoylhydrolase [Enterococcus hirae] pir A42296 A42296 lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus irae (ATCC 9790)	75	60
266	5	2224	1982	gnl PID e253211	ORF YDL065c [Saccharomyces cerevisiae]	75	50
269	2	1477	707	gi 1736647	ORF_ID:0347#4; similar to [SwissProt Accession Number P44634] [Escherichia coli]	75	61
276	11	7415	4593	gnl PID e221269	tail protein [Bacteriophage CP-1]	75	54
279	17	14992	14651	gi 1389549	ORF3 [Bacillus subtilis]	75	61
292	11	7829	8470	gi 160693	sporozoite surface protein [Plasmodium yoelii]	75	50
295	2	489	1157	gi 533099	endonuclease III [Bacillus subtilis]	75	59
307	4	3804	4889	gi 1321625	exo-alpha-1, 4-glucosidase [Bacillus stearothermophilus]	75	60
322	4	1088	1996	gi 310303	mosA [Rhizobium meliloti]	75	63
331	1	1	294	gi 1016092	ribosomal protein S14 [Cyanophora paradoxa]	75	57
334	7	6860	7969	gi 409286	bmrU [Bacillus subtilis]	75	45

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
340	1	3	743	gi 288413	glutamate dehydrogenase (NADP+) [Corynebacterium glutamicum] pir S32227 S32227 glutamate dehydrogenase (NADP+) (EC 1.4.1.4) - orynebacterium glutamicum	75	60
343	2	1497	778	gi 46602	putative transposase (AA 1 - 224) [Staphylococcus aureus] ir S12093 S12093 probable IS43mec protein - Staphylococcus aureus p P19380 TRA2_STAAU TRANSPOSASE FOR INSERTION SEQUENCE-LIKE ELEMENT 431MEC.	75	54
372	3	865	1629	gi 146282	gut operon repressor (gutR) [Escherichia coli]	75	58
372	7	6614	5307	gnl PID e255128	trigger factor [Bacillus subtilis]	75	62
387	3	1721	1353	gi 580902	ORF6 gene product [Bacillus subtilis]	75	53
399	30	28774	29805	gi 146278	glucitol-specific enzyme II (guta) [Escherichia coli] pir A26725 WQEC2S phosphotransferase system enzyme II (EC .7.1.69), sorbitol-specific, factor II - Escherichia coli sp P05705 PTHB_ECOLI PTS SYSTEM, GLUCITOL/SORBITOL-SPECIFIC IIBC COMPONENT (EIIBC-GUT)	75	61
399	33	31077	32768	gi 517205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	75	59
404	6	4994	4332	gi 1303921	YqiF [Bacillus subtilis]	75	64
404	7	4984	4829	gi 1303921	YqiF [Bacillus subtilis]	75	60
419	1	320	3	gi 496283	lysine [Bacteriophage Tuc2009]	75	67
431	3	1139	759	sp P46351 YZGD_BAC SU	HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5 REGION.	75	60
473	1	166	2	gnl PID e229299	R04D3.8 [Caenorhabditis elegans]	75	35
481	1	1	351	gi 1573766	phosphoglyceromutase (gpmA) [Haemophilus influenzae]	75	64
492	1	440	3	gi 806487	ORF211; putative [Lactococcus lactis]	75	57

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
595	1	705	181	gi 147485	queA [Escherichia coli]	75	51
619	2	879	319	gi 1063246	low homology to P14 protein of Haemophilus influenzae and 14.2 kDa protein of Escherichia coli [Bacillus subtilis]	75	59
663	1	15	1544	gi 475112	enzyme IIabc [Pedococcus pentosaceus]	75	54
701	4	662	946	gi 143793	tyrosyl-tRNA synthetase [Bacillus caldotenax]	75	60
719	1	970	419	gi 727436	putative 20-kDa protein [Lactococcus lactis]	75	56
886	1	101	409	gi 143150	levR [Bacillus subtilis]	75	59
939	1	403	191	gi 425467	transposase [Lactobacillus helveticus]	75	53
984	2	66	227	gi 1652190	Fat protein [Synechocystis sp.]	75	48
17	2	2592	2924	gi 532556	ORF23 [Enterococcus faecalis]	74	53
17	25	24449	25639	gi 1458228	mutY homolog [Homo sapiens]	74	50
21	7	4729	5229	gi 726320	putative protein of unknown function encoded by the IS200-like element [Yersinia pestis]	74	57
32	9	5819	4488	gi 1498962	M. jannaschii predicted coding region MJ0188 [Methanococcus jannaschii]	74	41
38	1	707	3	gi 142152	sulfate permease (gtg start codon) [Synechococcus PCC6301] pir A30301 GRYCS7 sulfate transport protein - Synechococcus sp. PCC 7942)	74	53
44	1	1	927	gi 1377823	aminopeptidase [Bacillus subtilis]	74	63
60	8	8747	8070	gi 143368	phosphoribosylformyl glycine synthetase I (PUR-L; gtg start'odon) [Bacillus subtilis]	74	63
72	8	7388	7119	gnl PID e209004	glutaredoxin-like protein [Lactococcus lactis]	74	53
91	4	1031	2257	gi 726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus subtilis]	74	58
105	7	5553	5855	gi 467418	unknown [Bacillus subtilis]	74	63



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
110	18	16903	15842	gi 45288	arcB (AA 1-336) [Pseudomonas aeruginosa]	74	57
112	3	1112	636	gi 887824	ORF_o310 [Escherichia coli]	74	53
123	8	6105	7619	gi 1773191	similar to Pseudomonas sp. ORF5, [Escherichia coli]	74	60
128	1	2	1315	gi 143961	pyruvate phosphate dikinase [Clostridium symbiosum] pir A36231 KIQAPO pyruvate,orthophosphate dikinase (EC 2.7.9.1) - lostridium symbiosum	74	58
128	26	18866	20401	gi 1303961	YqjJ [Bacillus subtilis]	74	57
150	5	4653	5303	gi 495046	tripeptidase [Lactococcus lactis]	74	53
159	8	7500	6850	gi 581098	GlnQ (AA 1-240); gtg start [Escherichia coli]	74	53
179	1	1259	57	gi 537080	ribonucleoside triphosphate reductase [Escherichia coli] pir A47331 A47331 oxygen-sensitive ribonucleoside-triphosphate eductase (EC 1.17.4.-) - Escherichia coli	74	62
183	2	1669	224	gi 1146200	DNA or RNA helicase, DNA-dependent ATPase [Bacillus subtilis]	74	53
213	4	2265	3200	gi 1373157	orf-X; hypothetical protein; Method: conceptual translation supplied by author [Bacillus subtilis]	74	63
229	13	13774	12806	gnl PID e290288	Met-tRNA <sup>I</sup> formyl transferase [Bacillus subtilis]	74	55
238	31	28648	28052	gi 451072	di-tripeptide transporter [Lactococcus lactis]	74	56
244	8	6409	5552	gi 467422	unknown [Bacillus subtilis]	74	60
249	1	7	411	gi 1591758	diaminopimelate epimerase [Methanococcus jannaschii]	74	51
270	3	1832	3955	gi 1303829	YqfK [Bacillus subtilis]	74	55
276	3	1668	1357	gi 496282	holin [Bacteriophage Tuc2009]	74	54
288	9	5807	5076	gi 530063	glycerol uptake facilitator [Streptococcus]	74	60

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
292	21	16780	17547	gi 1573646	pneumoniae] sp P52281 GLPF_STRPN GLYCEROL UPTAKE FACILITATOR PROTEIN.	74	42
297	1	682	11	gnl PID e255093	Mg(2+) transport ATPase protein C (mgtC) (SP:P22037) [Haemophilus influenzae]	74	54
298	3	3562	3095	gi 1303970	hypothetical protein [Bacillus subtilis]	74	46
321	10	5081	6028	pir A32950 A32950	YqjS [Bacillus subtilis]	74	56
327	2	904	3285	gi 1573876	probable reductase protein - Leishmania major	74	53
334	5	3942	5432	gi 1652678	virulence associated protein homolog (vacB) [Haemophilus influenzae]	74	57
341	13	13007	12069	gi 39881	amidase [Synecocystis sp.]	74	53
362	7	3529	5274	gnl PID e255093	ORF 311 (AA 1-311) [Bacillus subtilis]	74	58
376	3	1282	2346	gi 1773090	hypothetical protein [Bacillus subtilis]	74	59
421	2	48	1400	gi 1710632	transfer RNA-guanine transglycosylase [Escherichia coli]	74	58
471	1	815	3	gi 854234	beta-glucosidase [Bacillus subtilis]	74	53
480	2	263	607	gi 1303994	cymG gene product [Klebsiella oxytoca]	74	48
518	7	4409	5002	gi 145821	YqkM [Bacillus subtilis]	74	47
539	8	6607	7179	gi 1165295	EBG enzyme alpha subunit [Escherichia coli]	74	57
542	1	750	4	gi 1064810	D3703.8p [Saccharomyces cerevisiae]	74	56
559	1	1204	5	gi 43821	function unknown [Bacillus subtilis]	74	58
579	3	1373	1624	gi 1237013	nifJ protein (AA 1-1171) [Klebsiella pneumoniae] p P03833 NIFJ_KLEPN PYRUVATE-FLAVODOXIN OXIDOREDUCTASE (EC -.-.-).	74	46
624	4	2518	3669	gi 467394	ORF2 [Bacillus subtilis]	74	56
688	1	623	3	gi 62880	recombination protein [Bacillus subtilis]	74	48
763	1	106	441	gi 153955	novel hemolytic factor [Bacillus cereus]	74	46
811	1	3	158	gi 309662	envM protein [Salmonella typhimurium]	74	57
852	1	2	601	gi 309662	pheromone binding protein [Plasmid pCF10]	74	53

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
935	1	976	2	gi 467403	serY1-trRNA synthetase [Bacillus subtilis]	74	59
22	2	2178	2471	gi 467460	unknown [Bacillus subtilis]	73	61
24	2	1126	3150	gi 1303822	YqfF [Bacillus subtilis]	73	54
33	6	6638	6970	gi 536971	ORF_o76 [Escherichia coli]	73	56
48	1	621	1241	gnl PID e274111	aggregation promoting protein [Lactobacillus gasseri]	73	67
48	6	5327	7225	gi 1185289	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase [Bacillus subtilis]	73	56
50	2	1097	2008	gi 1498295	homoserine kinase homolog [Streptococcus pneumoniae]	73	55
52	4	2793	4334	gi 473902	alpha-acetolactate synthase [Lactococcus lactis]	73	59
55	1	1	261	gi 396365	alternate name yjba [Escherichia coli]	73	36
60	6	5935	5549	gi 551881	amidophosphoribosyltransferase [Lactobacillus casei] pir PC1136 PC1136 purF protein - Lactobacillus casei (fragment) sp P35853 PUR1_LACCA AMIDOPHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.14) GLUTAMINE PHOSPHORIBOSYLPYROPHOSPHATE AMIDOTRANSFERASE (ATASE) FRAGMENT	73	57
74	2	477	1355	gnl PID e233567	unknown [Mycobacterium tuberculosis]	73	54
81	19	14213	13845	gi 606073	ORF_ol69 [Escherichia coli]	73	52
93	7	2861	4075	gi 405134	acetate kinase [Bacillus subtilis]	73	56
100	1	1057	2	gi 1353561	ORF44 [Bacteriophage rlt]	73	52
100	41	28872	28627	gi 188492	heat shock-induced protein [Homo sapiens]	73	42
104	4	5558	5274	gi 312440	aspartate carbamoyltransferase [Bacillus caldolyticus] pir S34318 S34318 aspartate carbamoyltransferase (EC 2.1.3.2) - acillus caldolyticus	73	55
119	5	3264	3638	gi 473707	positive regulator for virulence factors [Clostridium perfringens]	73	39

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
123	17	16156	15665	gi 1303703	YrkD [Bacillus subtilis]	73	37
123	18	16133	16465	gi 1303893	YghL [Bacillus subtilis]	73	43
124	3	2165	1722	gi 486661	Tmm related protein [Saccharomyces cerevisiae]	73	45
127	6	5778	5101	gi 290561	ol88 [Escherichia coli]	73	48
128	10	6896	7201	pir S37387 S37387	internalin A precursor - Listeria monocytogenes	73	53
137	2	980	1954	gi 1276882	Epsi [Streptococcus thermophilus]	73	56
141	3	942	2777	gi 467336	unknown [Bacillus subtilis]	73	49
146	7	5611	4739	gi 149395	lacC [Lactococcus lactis]	73	56
154	6	3566	4621	gi 1354775	pfoS/R [Treponema pallidum]	73	46
155	8	7136	6726	gnl PID e247026	orf6 [Lactobacillus sake]	73	61
158	8	8693	7119	gi 1674275	(AE000056) Mycoplasma pneumoniae, hypothetical ABC transporter (yjcW) homolog; similar to Swiss-Prot Accession Number P32721, from E. coli [Mycoplasma pneumoniae]	73	45
162	4	4039	3305	gi 142997	glycerol uptake facilitator [Bacillus subtilis]	73	55
165	4	3962	3105	gi 882736	ORF_f278 [Escherichia coli]	73	58
171	3	3952	4689	gnl PID e63527	FtsE [Mycobacterium tuberculosis]	73	56
171	5	5673	6596	gi 1303854	YqgG [Bacillus subtilis]	73	59
179	9	9302	10414	gnl PID e254984	hypothetical protein [Bacillus subtilis]	73	55
180	1	24	1151	gi 43985	nifs-like gene [Lactobacillus delbrueckii]	73	56
181	12	10036	9674	gnl PID e220317	chorismate mutase [Staphylococcus xylosus]	73	50
181	13	10713	10003	gi 39813	phospho-2-dehydro-3-deoxyheptonate aldolase [Bacillus subtilis]	73	56
					ir S21418 S21418 phospho-2-dehydro-3-deoxyheptonate aldolase (EC 1.2.15) - Bacillus subtilis		
183	3	2716	1667	gi 1146199	putative [Bacillus subtilis]	73	36

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
198	1	869	108	gi 142854	homologous to <i>E. coli</i> radC gene product and to unidentified protein rom <i>Staphylococcus aureus</i> [Bacillus subtilis]	73	47
210	1	956	3	gnl PID e281310	acetyl coenzyme A acetyltransferase (thiolase) [Thermoanaerobacterium thermosaccharolyticum]	73	54
230	1	1	171	gi 304143	S-layer protein [Bacillus circulans]	73	46
235	1	715	2	gi 1732315	transport system permease homolog [Listeria monocytogenes]	73	49
235	2	888	676	gi 551726	sporulation protein [Bacillus subtilis]	73	54
242	4	3290	3517	gnl PID e236570	orf6 gene product [Enterococcus faecalis]	73	30
242	8	5914	6492	gi 1742340	HipB protein. [Escherichia coli]	73	49
250	3	3037	2411	gi 1174238	TipB [Pseudomonas fluorescens]	73	57
254	5	1124	792	gi 580900	ORF3 gene product [Bacillus subtilis]	73	52
269	9	5507	5154	gi 1303790	YqeI [Bacillus subtilis]	73	60
269	12	7989	7345	gi 285621	undefined open reading frame [Bacillus stearothermophilus]	73	54
284	1	1	915	gi 455528	ORF2 [Streptococcus thermophilus bacteriophage]	73	54
290	3	1932	2678	gnl PID e248883	unknown [Mycobacterium tuberculosis]	73	57
295	8	4521	4739	gi 145478	putative [Escherichia coli]	73	56
296	1	2	1846	gnl PID e249642	transketolase [Bacillus subtilis]	73	59
310	4	3488	3036	gi 1591900	nucleoside diphosphate kinase [Methanococcus jannaschii]	73	48
313	1	17	778	gi 1658371	cyclic beta-1,2-glucan modification protein [Rhizobium meliloti]	73	60
314	3	2642	2067	gi 1330343	C34D4.12 gene product [Caenorhabditis elegans]	73	56
325	1	492	4	gi 407908	EIIscr [Staphylococcus xylosus]	73	56
345	19	20549	21901	gi 443691	glutathione reductase [Streptococcus thermophilus]	73	59

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
359	4	3280	2252	gi 1001478	hypothetical protein [Synecocystis sp.]	73	50
374	1	884	3	gi 435123	PacL [Synecococcus sp.]	73	58
379	6	5676	4339	gi 887822	possible frameshift at end to join to next ORF? [Escherichia coli]	73	57
383	4	3815	3387	gi 1651732	mutator MutT protein [Synecocystis sp.]	73	52
392	4	3454	5202	gi 294587	minimal change nephritis transmembrane glycoprotein [Rattus orvegicus]	73	56
394	5	4267	5250	gi 49011	amidinotransferase II [Streptomyces griseus]	73	42
395	10	4252	4608	gi 1591139	M. jannaschii predicted coding region MJ0435 [Methanococcus jannaschii]	73	48
397	1	885	4	gnl PID e249658	GrlA [Bacillus subtilis]	73	56
399	15	10007	11569	gi 565619	citrate lyase alpha-subunit [Klebsiella pneumoniae] pir S60776 S60776 citrate (pro-3S)-lyase (EC 4.1.3.6) alpha chain - lebsiella pneumoniae	73	54
416	2	660	1649	gi 475114	regulatory protein [Pediococcus pentosaceus]	73	50
436	6	4124	3540	gi 727436	putative 20-kDa protein [Lactococcus lactis]	73	53
446	3	1618	4260	gi 882711	exonuclease V alpha-subunit [Escherichia coli]	73	48
462	1	819	43	gi 1399011	immunogenic secreted protein precursor [Streptococcus pyogenes]	73	63
482	5	3181	2501	gi 1072419	glcB gene product [Staphylococcus carnosus]	73	55
495	4	1340	3031	gi 146547	kdpA [Escherichia coli]	73	55
523	4	2354	1821	pir A00392 RDSODF	dihydrofolate reductase (EC 1.5.1.3) - Enterococcus faecium	73	54
543	5	3099	2893	gi 19743	nsGRP-2 [Nicotiana sylvestris]	73	53
567	1	9	740	gi 1147601	cyclophilin isoform 4 [Caenorhabditis elegans]	73	54

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
629	1	945	4	gi 1006620	ABC transporter [Synechocystis sp.]	73	46
714	2	344	556	gi 1045872	ATP-binding protein [Mycoplasma genitalium]	73	61
747	1	320	3	gi 437389	transposase [Lactococcus lactis]	73	56
764	1	3	515	gi 532554	ORF21 [Enterococcus faecalis]	73	50
766	1	683	3	gi 1673788	(AE000015) Mycoplasma pneumoniae, fructose-bisphosphate aldolase; similar to Swiss-Prot Accession Number P13243, from <i>B. subtilis</i> [Mycoplasma pneumoniae]	73	52
880	1	198	4	gi 309661	regulatory protein [Plasmid pCF10]	73	50
897	1	3	170	gi 807976	unknown [Saccharomyces cerevisiae]	73	57
5	1	223	2	gnl PID e255315	unknown [Mycobacterium tuberculosis]	72	56
8	5	4158	4799	gi 587088	shikimate kinase [Bacillus subtilis]	72	54
19	6	2600	2833	gi 34844	embryonic myosin heavy chain (AA 1 - 1940) [Homo sapiens] ir S04090 S04090 myosin heavy chain, skeletal muscle, embryonic - man	72	38
19	25	12872	14605	gnl PID e242896	orf5 [Bacteriophage A2]	72	52
21	4	2777	2598	gi 54115	skeletal muscle chloride channel [Mus musculus domesticus]	72	45
23	7	3702	4847	gi 144714	NADPH-dependent butanol dehydrogenase [Clostridium acetobutylicum] pir JU0053 JU0053 NADPH-dependent butanol dehydrogenase - lostridium acetobutylicum	72	48
32	1	1073	3	gi 1303839	YqfR [Bacillus subtilis]	72	50
39	8	4137	3244	pir A32950 A32950	probable reductase protein - Leishmania major	72	55
43	3	969	1919	gi 290494	o287 [Escherichia coli]	72	46
45	2	911	1567	gi 1039479	ORFU [Lactococcus lactis]	72	50
55	6	2549	2896	gi 755602	unknown [Bacillus subtilis]	72	51
55	7	3178	3660	gi 1303914	YqHY [Bacillus subtilis]	72	49

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
60	1	1302	34	gi 143374	phosphoribosyl glycine synthetase (PUR-D; gtg start codon) <i>Bacillus subtilis</i>	72	59
60	3	3422	2838	gi 143372	phosphoribosyl glycine synthetase (PUR-N) <i>Bacillus subtilis</i>	72	48
60	10	9771	9010	gi 143367	phosphoribosyl aminidazole succinocarboxamide synthetase (PUR-C; tg start codon) <i>Bacillus subtilis</i>	72	57
70	5	3615	3833	sp P43672 YCBH_ECO LI	HYPOTHETICAL 14.4 KD PROTEIN IN PYRD-PQIA INTERGENIC REGION.	72	48
79	2	632	841	gi 1652343	ABC transporter [ <i>Synechocystis</i> sp.]	72	47
85	2	1843	770	gi 1354775	pfoS/R [ <i>Treponema pallidum</i> ]	72	45
87	1	2	745	gi 42029	ORF1 gene product [ <i>Escherichia coli</i> ]	72	47
88	1	124	1047	gi 535348	CodV [ <i>Bacillus subtilis</i> ]	72	50
88	7	3862	4752	gi 149413	ORF [ <i>Lactococcus lactis</i> ]	72	51
91	2	611	877	gi 726480	L-glutamine-D-fructose-6-phosphate amidotransferase [ <i>Bacillus subtilis</i> ]	72	57
98	16	16302	15163	gi 147326	transport protein [ <i>Escherichia coli</i> ]	72	57
101	6	4676	4023	gi 1109685	ProW [ <i>Bacillus subtilis</i> ]	72	53
104	3	5331	3982	gi 312441	dihydroorotase [ <i>Bacillus caldolyticus</i> ]	72	58
114	10	11165	12205	gi 556881	Similar to <i>Saccharomyces cerevisiae</i> SUA5 protein [ <i>Bacillus subtilis</i> ] pir S49358 S49358 ipc-29d protein - <i>Bacillus subtilis</i> sp P39153 YWLC_BACSU HYPOTHETICAL 37.0 KD PROTEIN IN SPOIR-GLYC NTERGENIC REGION.	72	60
128	19	14325	11560	gi 143150	levR [ <i>Bacillus subtilis</i> ]	72	58
130	2	382	1437	gi 308850	ATP binding protein [ <i>Lactococcus lactis</i> ]	72	55
135	4	5012	3693	gi 413940	ipa-16d gene product [ <i>Bacillus subtilis</i> ]	72	56
150	6	5114	5878	gi 495046	tripeptidase [ <i>Lactococcus lactis</i> ]	72	54
154	9	5850	5677	gi 425467	transposase [ <i>Lactobacillus helveticus</i> ]	72	52



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
168	4	1375	1563	gi 1652869	NADH dehydrogenase [Synechocystis sp.]	72	55
173	5	2879	4024	gnl PID e254877	unknown [Mycobacterium tuberculosis]	72	57
179	2	1608	2399	gi 709993	hypothetical protein [Bacillus subtilis]	72	45
179	6	7584	7844	gi 1161934	DltC [Lactobacillus casei]	72	54
180	21	19948	21105	gi 1773197	similar to M. fervidus malate dehydrogenase [Escherichia coli]	72	55
182	1	3	413	gi 1146182	putative [Bacillus subtilis]	72	48
200	23	13106	12789	gi 1707358	polyprotein precursor [Soybean mosaic virus]	72	34
204	6	2462	2289	gi 1200525	dihydrolipoamide acetyltransferase [Pseudomonas aeruginosa]	72	61
204	9	6374	5187	gi 1732040	alcohol dehydrogenase [Actinobacillus pleuropneumoniae]	72	56
205	1	463	71	gi 42029	ORF1 gene product [Escherichia coli]	72	57
210	7	6433	5279	gi 142978	glycerol dehydrogenase [Bacillus stearothermophilus] pir JQ1474 JQ1474	72	46
					glycerol dehydrogenase (EC 1.1.1.6) - Bacillus tearothermophilus		
213	6	4086	5141	gi 431231	uracil permease [Bacillus caldolyticus]	72	51
223	1	99	833	gi 1573615	ATP-binding protein (abc) [Haemophilus influenzae]	72	47
227	1	26	886	gi 1070015	protein-dependent [Bacillus subtilis]	72	52
228	4	2047	2481	gi 467339	unknown [Bacillus subtilis]	72	50
238	17	14728	15582	gi 882736	ORF_f278 [Escherichia coli]	72	59
250	6	4169	4765	gi 437389	transposase [Lactococcus lactis]	72	56
258	7	5296	7089	gi 192185	acid beta-galactosidase [Mus musculus]	72	53
266	3	2024	1773	gi 145149	ORF [Escherichia coli]	72	50
269	8	5142	4477	gi 1303791	YqeJ [Bacillus subtilis]	72	45
276	13	9843	8152	gnl PID e59644	predicted 86.4kd protein; 52Kd observed [Mycobacteriophage 15]	72	48
278	2	965	1573	gi 425467	transposase [Lactobacillus helveticus]	72	52

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
279	2	1305	340	gnl PID e198981	ttg_start [Campylobacter coli]	72	47
283	4	1668	2045	gi 1353563	ORF46 [Bacteriophage rlt]	72	48
286	2	789	2606	gi 1651216	P2-peptidase [Bacillus licheniformis]	72	52
290	4	2676	3239	gi 1653645	ribosome releasing factor [Synechocystis sp.]	72	56
301	2	1762	899	gi 606013	CG Site No. 829 [Escherichia coli]	72	57
362	2	377	688	gi 1001826	cadmium-transporting ATPase [Synechocystis sp.]	72	53
369	1	582	142	gi 153745	mannitol-specific enzyme III [Streptococcus mutans] pir B44798 B44798 mannitol-specific factor III, MtlF - treptococcus mutans	72	47
379	2	1934	1527	gi 1055071	C23G10.2 gene product [Caenorhabditis elegans]	72	51
384	2	694	1098	gi 1208474	hypothetical protein [Synechocystis sp.] (AE000018) Mycoplasma pneumoniae, osmotically inducible protein; similar to Swiss-Prot Accession Number P23929, from E. coli [Mycoplasma pneumoniae]	72	49
388	1	291	4	gi 1673836		72	43
401	6	3995	5137	gi 508242	ORF 6, putative Galf synthesis pathway protein [Escherichia coli] gi 510253.orf6 [Escherichia coli]	72	62
404	2	2119	776	gi 466474	cellobiose phosphotransferase enzyme II'' [Bacillus tearothermophilus]	72	48
416	4	3461	1980	gi 710632	beta-glucosidase [Bacillus subtilis]	72	55
416	7	6285	5551	gnl PID e269549	Unknown [Bacillus subtilis]	72	52
419	3	759	505	gi 928830	ORF75; putative [Lactococcus lactis] phage BK5-T]	72	47
441	4	3420	4676	gi 1732195	beta-cystathionase [Vibrio furnissii]	72	54
460	3	1385	2641	gi 1652389	beta ketoacyl-acyl carrier protein synthase [Synechocystis sp.]	72	55
460	5	3129	3560	gnl PID e289141	similar to hydroxymyristoyl-acyl carrier	72	54

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
460	8	5817	6023	gi 285621	protein) dehydratase [Bacillus subtilis]		
462	2	1591	785	gi 148304	undefined open reading frame [Bacillus stearothermophilus]	72	57
467	1	2	706	gi 148711	beta-1,4-N-acetylmuramoylhydrolase [Enterococcus hirae] pir A42296 A42296 lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus irae (ATCC 9790)	72	51
469	3	1144	1419	gi 466474	6-aminohexanoate-cyclic-dimer hydrolase [Flavobacterium sp.] gi 488343 6-aminohexanoate-cyclic-dimer hydrolase [Flavobacterium p.]	72	50
493	1	1124	240	sp P50848 YPWA_BAC SU	cellobiose phosphotransferase enzyme II'' [Bacillus tearothermophilus]	72	48
536	2	379	218	gi 437389	HYPOTHETICAL 58.2 KD PROTEIN IN KDGT-XPT INTERGENIC REGION.	72	58
543	1	574	86	gi 290513	transposase [Lactococcus lactis]	72	47
592	1	57	680	gi 987092	f470 [Escherichia coli]	72	55
666	2	551	967	gi 1064786	ABC-transporter [Streptomyces hygroscopicus]	72	48
762	1	974	273	gi 304928	function unknown [Bacillus subtilis]	72	55
792	1	401	3	pir A36933 A36933	pantothenate synthetase [Escherichia coli]	72	50
873	1	183	4	gnl PID e258329	diacylglycerol kinase homolog - Streptococcus mutans	72	55
4	4	3799	3155	gi 496943	oxaloacetate decarboxylase alpha-chain [Legionella pneumophila]	71	45
10	2	180	977	gnl PID e234078	ORF [Saccharomyces cerevisiae]	71	49
16	7	4922	6097	gi 534982	hom [Lactococcus lactis]	71	54
21	6	4148	3972	gi 1736645	phosphoglucosyltransferase [Spinacia oleracea]	71	50
23	27	16452	17459	gi 1408503	Proline/betaine transporter (Proline porter II) (PPII). [Escherichia coli]	71	52
25	7	5812	6669	gi 413943	yeR gene product [Bacillus subtilis]	71	58

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
31	1	80	946	gi 534045	antiterminator [Bacillus subtilis]	71	47
39	3	755	1297	sp P09997 VIDA_ECO LI	HYPOTHETICAL 29.7 KD PROTEIN IN IBPA-GYRB INTERGENIC REGION.	71	50
39	7	2537	3193	pir C43748 C43748	hypothetical protein (pepX 3' region) - Lactococcus lactis subsp. lactis	71	54
45	10	5119	5484	gi 606044	ORF_ol30; Geneplot suggests frameshift, none found [Escherichia coli]	71	51
48	10	11722	10148	gi 20432	4-coumarate:CoA ligase Pc4Cl-1 (AA 1-544) [Petroselinum crispum] ir S01667 S01667 4-coumarate--CoA ligase (EC 6.2.1.12) (clone 4CL-1) - parsley	71	39
55	4	1470	1709	gi 1303901	YqhT [Bacillus subtilis]	71	54
57	10	12899	13060	gi 40053	phenylalanyl-tRNA synthetase alpha subunit [Bacillus subtilis] ir S11730 YFBSA phenylalanine--tRNA ligase (EC 6.1.1.20) alpha ain - Bacillus subtilis	71	45
58	3	3743	2571	gi 1658403	formate dehydrogenase alpha subunit [Moorella thermoacetica]	71	51
68	11	8225	8602	gi 793910	surface antigen [Homo sapiens]	71	49
74	4	2908	2042	gi 467435	unknown [Bacillus subtilis]	71	55
85	3	3267	1966	gi 142613	branched chain alpha-keto acid dehydrogenase E2 [Bacillus subtilis] gi 1303944 BfmBB [Bacillus subtilis]	71	56
111	8	5737	4253	gi 1256135	YbbF [Bacillus subtilis]	71	50
111	9	6590	5730	gi 1573762	glucokinase regulator [Haemophilus influenzae]	71	53
120	1	111	353	gnl PID e235823	unknown [Schizosaccharomyces pombe]	71	52
123	11	10387	11196	gi 1773195	hypothetical [Escherichia coli]	71	55
151	3	4045	3098	gi 1256618	transport protein [Bacillus subtilis]	71	51
172	6	3949	4806	gi 1262288	CdsA [Brucella abortus]	71	56
172	7	5264	6448	gi 40100	rodC (tag3) polypeptide (AA 1-746) [Bacillus subtilis] ir S06049 S06049 rodC	71	52

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					protein - <i>Bacillus subtilis</i> p P13485 TAGF_BACSU TEICHOIC ACID BIOSYNTHESIS PROTEIN F.		
190	7	3454	3122	gi 532556	ORF23 [ <i>Enterococcus faecalis</i> ]	71	52
195	24	9850	11871	gi 405564	traE [Plasmid pSK41]	71	45
215	4	3361	2711	gi 1573086	uridine kinase (uridine monophosphokinase) (udk) [ <i>Haemophilus influenzae</i> ]	71	51
218	2	1456	2613	gnl PID e254644	membrane protein [ <i>Streptococcus pneumoniae</i> ]	71	41
222	3	1205	2053	gnl PID e255114	glutamate racemase [ <i>Bacillus subtilis</i> ]	71	56
222	4	1611	1387	gi 1001195	phosphate transport system permease protein PstA [ <i>Synechocystis</i> sp.]	71	57
222	14	8852	9853	gi 466720	No definition line found [ <i>Escherichia coli</i> ]	71	53
238	22	19256	20578	gi 595299	Ygik [ <i>Salmonella typhimurium</i> ]	71	50
255	3	2692	1061	gnl PID e254877	unknown [ <i>Mycobacterium tuberculosis</i> ]	71	55
265	5	2960	1581	gi 1039479	ORFU [ <i>Lactococcus lactis</i> ]	71	58
276	2	1359	538	gi 496283	lysine [ <i>Bacteriophage Tuc2009</i> ]	71	63
290	5	3552	4379	gi 1016162	ABC transporter subunit [ <i>Cyanophora paradoxa</i> ]	71	49
290	7	5659	6912	gi 1001708	Nifs [ <i>Synechocystis</i> sp.]	71	56
292	3	948	2156	gnl PID e233874	hypothetical protein [ <i>Bacillus subtilis</i> ]	71	55
318	4	3229	2285	gi 1256138	YbbI [ <i>Bacillus subtilis</i> ]	71	54
333	1	145	741	gi 293011	unknown protein [ <i>Lactococcus lactis</i> ]	71	50
344	1	76	396	gi 853775	unknown [ <i>Bacillus subtilis</i> ]	71	53
350	1	138	1394	gi 1652389	beta ketoacyl-acyl carrier protein synthase [ <i>Synechocystis</i> sp.]	71	57
363	4	4184	5674	gi 1657518	similar to fdrA gene of <i>E. coli</i> [ <i>Escherichia coli</i> ]	71	54
364	5	5319	6563	gi 1657522	hypothetical protein [ <i>Escherichia coli</i> ]	71	46
367	13	6539	6162	gi 44225	ribosomal protein L18 (AA 1-116)	71	51

Table 3: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					[Mycoplasma capricolum] ir S02847 R5YM18 ribosomal protein L18 - Mycoplasma capricolum GC3)		
379	7	6884	5655	gi 887821	ORF_0398 [Escherichia coli]	71	50
399	9	6528	7664	gi 154198	oxaloacetate decarboxylase [Salmonella typhimurium] pir C44465 C44465 sodium ion pump oxaloacetate decarboxylase ubunit beta - Salmonella typhimurium	71	50
399	18	13540	14778	gi 143165	malic enzyme (EC 1.1.1.38) [Bacillus stearothermophilus] pir A33307 DEBSXS malate dehydrogenase oxaloacetate-decarboxylating) (EC 1.1.1.38) - Bacillus stearothermophilus	71	46
404	4	3769	3029	gi 143402	recombination protein (ttg start codon) [Bacillus subtilis] gi 1303923 RecN [Bacillus subtilis]	71	48
464	1	1532	216	gi 895749	putative cellobiose phosphotransferase enzyme II'' [Bacillus subtilis]	71	40
464	3	2088	2846	gi 1486242	unknown [Bacillus subtilis]	71	39
481	2	954	409	gi 144729	butanol dehydrogenase [Clostridium acetobutylicum] sp Q04944 ADHA_CLOAB NADH-DEPENDENT BUTANOL DEHYDROGENASE A (EC 1.1.1.-) (BDH I).	71	58
482	4	2503	1841	gi 1072418	glcA gene product [Staphylococcus carnosus]	71	58
496	2	1636	848	gi 1001226	methionine aminopeptidase [Synechocystis sp.]	71	51
503	2	1624	650	gi 39478	ATP binding protein of transport ATPases [Bacillus firmus] ir S15486 S15486 ATP-binding protein - Bacillus firmus p P26946 YATR_BACFI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING OTEIN.	71	49

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
513	2	1590	982	gnl PID e202290	unknown [Lactobacillus sake]	71	46
530	1	2	1534	gi 1542974	AbcA [Thermoanaerobacterium thermosulfurigenes]	71	52
537	1	706	365	gi 929972	ORFB; similar to B. anthracis SterneL element ORFB; putative S150-like transposase [Bacillus anthracis]	71	57
553	1	304	1287	gi 1653479	regulatory components of sensory transduction system [Synechocystis sp.]	71	48
573	9	5560	5090	gi 143799	MtrA [Bacillus subtilis]	71	59
583	1	21	341	gi 1064791	function unknown [Bacillus subtilis]	71	50
584	2	638	276	gi 662792	single-stranded DNA binding protein [unidentified eubacterium]	71	58
585	1	282	809	gi 666972	ORF 168 [Synechococcus sp.]	71	46
611	1	985	2	gi 1039479	ORFU [Lactococcus lactis]	71	55
616	1	350	3	gi 1088272	nitrogen fixation protein [Bacillus cereus]	71	52
624	1	61	399	gi 40014	pot. ORF 446 (aa 1-446) [Bacillus subtilis]	71	53
624	2	608	1732	gi 40015	pot. ORF 378 (aa 1-378) [Bacillus subtilis]	71	51
659	1	76	582	gi 1591045	hypothetical protein (SP:P31466) [Methanococcus jannaschii]	71	51
668	2	836	1030	gi 467330	replicative DNA helicase [Bacillus subtilis]	71	60
683	1	582	118	gnl PID e264663	CinA [Streptococcus pneumoniae]	71	55
701	3	411	797	gi 143795	transfer RNA-Tyr synthetase [Bacillus subtilis]	71	51
720	1	1	351	gi 1595810	type-I signal peptidase SpsB [Staphylococcus aureus]	71	55
724	2	1020	415	gnl PID e239621	ORF YNL218w [Saccharomyces cerevisiae]	71	51
790	2	658	383	gi 1783253	homologous to many ATP-binding transport proteins; hypothetical [Bacillus subtilis]	71	48

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
799	1	505	906	gi 580866	ipa-12d gene product [Bacillus subtilis]	71	45
974	2	139	333	gi 1778531	HI0021 homolog [Escherichia coli]	71	42
980	1	156	497	gi 437389	transposase [Lactococcus lactis]	71	57
4	3	3170	2418	gi 1001805	hypothetical protein [Synechocystis sp.]	70	55
17	21	18642	21527	gi 145821	EBG enzyme alpha subunit [Escherichia coli]	70	53
19	8	2894	3952	gi 1353527	ORF10 [Bacteriophage r1t]	70	58
23	6	2640	3230	gi 699336	C. freundii orfw homologue [Mycobacterium leprae] sp P53523 Y02Y_MYCLE HYPOTHETICAL 20.9 KD PROTEIN U471A.	70	43
27	3	1011	493	gi 1001644	regulatory components of sensory transduction system [Synechocystis sp.]	70	44
31	2	1095	1337	gi 1100076	PTS-dependent enzyme II [Clostridium longisporum]	70	55
32	10	6527	5817	gi 1591789	M. jannaschii predicted coding region MJ1163 [Methanococcus jannaschii]	70	51
33	7	6930	7235	gi 536972	ORF_090a [Escherichia coli]	70	45
35	2	500	2533	gi 43819	nagE gene product [Klebsiella pneumoniae]	70	50
47	13	15837	14512	gi 150209	ORF 1 [Mycoplasma mycoides]	70	44
49	15	10409	11179	gi 853751	N-acetylmuramoyl-L-alanine amidase [Bacteriophage A511]	70	54
57	7	8365	12189	gi 142440	ATP-dependent nuclease [Bacillus subtilis]	70	48
57	16	18656	18033	gi 388565	major cell-binding factor [Campylobacter jejuni]	70	52
59	9	4985	7060	gnl PID e254877	unknown [Mycobacterium tuberculosis]	70	49
72	6	6771	4600	gi 557567	ribonucleotide reductase R1 subunit [Mycobacterium tuberculosis] sp P50540 R1R1_MYCTU RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE ALPHA HAIN (EC 1.17.4.1) (RIBONUCLEOTIDE REDUCTASE) (R1 SUBUNIT) FRAGMENT).	70	53
76	8	5960	6343	gi 1063251	no homologous protein [Bacillus subtilis]	70	52



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
81	16	12529	11723	gi 1732200	PTS permease for mannose subunit IIPMan [Vibrio furnissii]	70	52
98	7	8974	7874	gi 1573045	hypothetical [Haemophilus influenzae]	70	46
110	2	1353	502	gi 1399848	unknown [Synecoccus PCC7942]	70	52
123	7	5009	5527	gi 143284	negative regulator pai 1 [Bacillus subtilis]	70	51
123	22	19729	20412	gi 1591493	glutamine transport ATP-binding protein Q [Methanococcus jannaschii]	70	48
133	6	5905	6498	gi 746399	transcription elongation factor [Escherichia coli]	70	50
134	1	1	384	gi 1146242	aspartate 1-decarboxylase [Bacillus subtilis]	70	49
138	10	8543	7953	gi 467371	LACI family of transcriptional repressor (probable) [Bacillus subtilis]	70	50
160	3	1263	1520	gi 1468939	meso-2,3-butanediol dehydrogenase (D-acetoin forming) [Klebsiella pneumoniae]	70	45
174	3	2279	1572	gi 413931	ipa-7d gene product [Bacillus subtilis]	70	44
177	2	2104	1022	gnl PID ei86242	D-mannonate hydrolase [Thermotoga neapolitana]	70	52
178	2	1320	532	gi 499659	K+ channel protein [Panulirus interruptus]	70	51
180	18	17770	18729	gi 887824	ORF_o310 [Escherichia coli]	70	50
180	22	21072	22526	gi 1573294	hypothetical [Haemophilus influenzae]	70	40
181	9	7409	6279	sp P20692 TYRA_BAC SU	PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH)	70	49
197	5	4529	6340	gi 1783252	homologous to many ATP-binding transport proteins including SwissProt:CYDD_ECOLI; hypothetical [Bacillus subtilis]	70	47
200	21	12419	11820	gi 290943	HindIII modification methyltransferase [Haemophilus influenzae] sp P43871 MTH3_HAEIN MODIFICATION METHYLASE HINDIII (EC 2.1.1.72) ADENINE-SPECIFIC METHYLTRANSFERASE HINDIII	70	47

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					(M.HINDIII).		
210	4	3877	3269	gi 602683	orfC [Mycoplasma capricolum]	70	47
217	2	405	707	gi 153767	ORF [Streptococcus pneumoniae]	70	56
222	8	4940	6046	gi 537033	ORF_f356 [Escherichia coli]	70	54
222	15	9825	10553	gi 537039	ORF_o228a [Escherichia coli]	70	56
227	4	1871	2893	gi 1070014	protein-dependent [Bacillus subtilis]	70	44
228	2	1343	792	gi 1742730	Protein AraJ precursor. [Escherichia coli]	70	50
228	5	3470	2574	gi 1573390	hypothetical [Haemophilus influenzae]	70	54
231	2	2470	1238	gi 1574085	H. influenzae predicted coding region HI1048 [Haemophilus influenzae]	70	48
235	4	2779	2138	gi 309662	pheromone binding protein [Plasmid pCF10]	70	46
239	4	5807	6409	gi 682765	mccB gene product [Escherichia coli]	70	41
248	1	3	350	gi 143725	putative [Bacillus subtilis]	70	52
254	4	838	497	gi 49318	ORF4 gene product [Bacillus subtilis]	70	48
256	3	1737	2612	gi 596092	putative multiple membrane domain protein; possible TTG initiation odon at position 1064, near putative RBS at position 1052 Streptococcus pyogenes]	70	51
279	15	14547	14224	gi 1389549	ORF3 [Bacillus subtilis]	70	50
283	6	2279	3190	gi 853751	N-acetylmuramoyl-L-alanine amidase [Bacteriophage A511]	70	52
292	8	5557	6534	gi 474195	This ORF is homologous to a 40.0 kd hypothetical protein in the htrB ' region from E. coli, Accession Number X61000 [Mycoplasma-like rganism]	70	50
294	8	2776	3375	gi 1750126	YncB [Bacillus subtilis]	70	47
294	10	3742	4020	gi 984581	YafQ [Escherichia coli]	70	50
299	1	905	132	gi 606309	ORF_o265; gtg start [Escherichia coli]	70	40
300	3	3200	2784	gi 289260	comE ORF1 [Bacillus subtilis]	70	50
301	9	8564	7590	gi 1303865	YggR [Bacillus subtilis]	70	52
336	2	661	921	gi 202864	[Rat alternatively spliced mRNA.], gene	70	47

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
339	1	269	3	gi 786163	product [Rattus norvegicus]		
351	9	4760	4359	gi 799235	Ribosomal Protein L10 [Bacillus subtilis]	70	50
399	28	28203	28793	gi 146278	dTDP-6-deoxy-L-lyxo-4-hexulose reductase [Escherichia coli]	70	45
406	1	1	552	gi 49315	glucitol-specific enzyme II (gutA) [Escherichia coli] pir A26725 WQEC2S	70	52
436	5	2417	2193	gi 773665	phosphotransferase system enzyme II (EC .7.1.69), sorbitol-specific, factor II - Escherichia coli sp P05705 PTHB_ECOLI PTS SYSTEM, GLUCITOL/SORBITOL-SPECIFIC IIBC OMPOENT (EIIBC-GUT)		
482	3	1887	1660	gi 48680	ORF1 gene product [Bacillus subtilis]	70	50
529	3	6587	7030	gi 1022726	transposase [Lactococcus lactis]	70	36
535	2	1702	965	gi 1747435	ptsG-like product [Bacillus subtilis]	70	47
543	2	1248	547	gi 1591045	unknown [Staphylococcus haemolyticus]	70	44
543	8	4084	3878	gi 511976	KdpE [Clostridium acetobutylicum]	70	52
560	3	1037	876	gi 558458	hypothetical protein (SP:P31466) [Methanococcus jannaschii]	70	47
573	4	1920	2258	gi 336639	SERP gene product [Plasmodium falciparum]	70	60
599	2	244	573	gi 42029	acidic 82 kDa protein [Homo sapiens]	70	40
608	3	867	556	gi 475032	prephytoene pyrophosphate dehydrogenase [Cyanophora paradoxa] gi 1016130 prenyl transferase [Cyanophora paradoxa] pir A40433 A40433 prephytoene pyrophosphatase dehydrogenase (crtE) omolog - Cyanophora paradoxa	70	32
					ORF1 gene product [Escherichia coli]	70	49
					formamidopyrimidine-DNA glycosylase [Streptococcus mutans] sp P55045 FPG_STRMU FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE (EC .2.2.23) (FAPY-DNA GLYCOSYLASE).	70	53

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
636	1	2	628	gi 606309	ORF_o265; gtg start [Escherichia coli]	70	50
670	2	2157	1828	gi 1657698	hyaluronan receptor [Homo sapiens]	70	41
702	1	103	870	gi 149490	sucrose-6-phosphate hydrolase [Lactococcus lactis] pir JH0754 JH0754 sucrose-6-phosphate hydrolase (EC 3.2.1.-) - actococcus lactis	70	51
726	2	725	480	gnl PID e240103	unknown ORF [Saccharomyces cerevisiae]	70	41
854	1	1	207	gi 532653	thermonuclease [Staphylococcus hyicus]	70	51
901	1	238	447	gi 172022	myosin 1 isoform (MYO2) [Saccharomyces cerevisiae]	70	20
940	1	1	318	gi 1039479	ORFU [Lactococcus lactis]	70	56
1	2	2112	1213	gi 413976	ipa-52r gene product [Bacillus subtilis]	69	51
8	2	2196	778	gi 1510108	ORF-1 [Agrobacterium tumefaciens]	69	50
8	9	7949	6654	gi 1196907	daunorubicin resistance protein [Streptomyces peucetius]	69	44
16	3	1618	2574	gi 1109684	proV [Bacillus subtilis]	69	53
17	26	25781	26944	gi 485275	53.6 kDa protein [Streptococcus pneumoniae]	69	44
17	35	32300	32770	gi 1574146	pfs protein (pfs) [Haemophilus influenzae]	69	53
23	30	18107	18538	gnl PID e249656	Ynet [Bacillus subtilis]	69	59
25	8	6653	6994	gi 413943	ipa-19d gene product [Bacillus subtilis]	69	46
37	2	2042	186	gi 143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir A27650 A27650 regulatory protein phoR - Bacillus subtilis sp P23545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-)	69	52
39	2	528	767	gi 1408493	homologous to SwissProt:YIDA_ECOLI hypothetical protein [Bacillus subtilis]	69	52
56	6	4809	3457	gi 1591610	probable ATP-dependent helicase [Methanococcus jannaschii]	69	45
67	5	3042	3938	gi 1658188	oxidative stress transcriptional regulator	69	39

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
68	3	684	1529	gnl PID e214719	[ <i>Erwinia carotovora</i> ]		
72	4	2099	3394	gi 882672	PlcR protein [ <i>Bacillus thuringiensis</i> ]	69	45
81	15	11820	10915	gi 1732201	ORF o313 [ <i>Escherichia coli</i> ]	69	37
83	20	14001	15800	gi 1230668	PTS permease for mannose subunit IIBMan [ <i>Vibrio furnissii</i> ]	69	44
85	6	6309	5299	sp P54533 DLD2_BAC SU	Similar to Arginyl-tRNA synthetase (Swiss Prot. accession number P11875) [ <i>Saccharomyces cerevisiae</i> ]	69	44
86	3	2084	3367	gi 143318	LIPAMIDE DEHYDROGENASE COMPONENT (E3) OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1.8.1.4) (DIHYDROLIPOAMIDE DEHYDROGENASE) (LPD-VAL)	69	46
94	2	1401	751	gi 755216	phosphoglycerate kinase [ <i>Bacillus megaterium</i> ]	69	53
94	16	20498	19197	gi 1208948	N-acetylmuramidase [ <i>Lactococcus lactis</i> ]	69	41
98	8	10201	9029	gi 563934	unknown [ <i>Escherichia coli</i> ]	69	47
109	4	2350	1316	gi 396501	similar to E. coli hypothetical protein: PIR Accession Number Q0614 [ <i>Bacillus subtilis</i> ]	69	51
114	1	83	1522	gi 1658402	aspartyl-tRNA synthetase [ <i>Thermus aquaticus thermophilus</i> ] pir S33743 S33743 aspartate--tRNA ligase (EC 6.1.1.12) - <i>Thermus quaticus</i>	69	56
123	9	7617	8984	gi 1773192	formate dehydrogenase beta subunit [ <i>Moorella thermoacetica</i> ]	69	45
128	11	7940	7578	gi 895750	similar to S. cerevisiae dal1 [ <i>Escherichia coli</i> ]	69	50
130	10	8764	9036	gi 1641	putative cellobiose phosphotransferase enzyme III [ <i>Bacillus ubtilis</i> ]	69	53
					put. Na(+)/glucose co-transporter (AA 1-662) [ <i>Oryctolagus cuniculus</i> ]  1717	69	47

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
138	26	16721	17545	pir A25805 A25805	cortical sodium-D-glucose cotransporter [Oryctolagus iculus]		
139	2	310	1083	gi 1408587	L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus subtilis	69	55
139	9	5196	4984	gi 473955	relaxase [Lactococcus lactis lactis]	69	46
142	9	5559	4564	gi 623073	DNA-binding protein [Lactobacillus sp.]	69	34
155	6	4658	5818	gi 1591260	ORF360; putative [Bacteriophage LL-H]	69	47
158	12	11671	11201	gi 606744	endoglucanase [Methanococcus jannaschii]	69	48
162	5	5888	4032	gi 142993	cytidine deaminase [Bacillus subtilis]	69	52
180	2	1901	1203	gi 1575577	glycerol-3-phosphate dehydrogenase (glpD) (EC 1.1.99.5) [Bacillus ubtilis]	69	54
197	4	3571	4602	gi 1783251	DNA-binding response regulator [Thermotoga maritima]	69	49
197	6	6283	7701	gi 1783253	homologous to cytochrome d ubiquinol oxidase subunit II; hypothetical [Bacillus subtilis]	69	46
222	1	201	10	gi 149901	homologous to many ATP-binding transport proteins; hypothetical [Bacillus subtilis]	69	49
223	28	23857	24567	gnl PID e269548	gene codes for a 19 kDa protein [Mycobacterium avium] sp P46733 19KD_MYCAV 19 KD LIPOPROTEIN ANTIGEN PRECURSOR.	69	50
228	3	2031	1285	gi 1742730	Unknown [Bacillus subtilis]	69	53
229	8	7390	6698	gi 1162980	Protein AraJ precursor. [Escherichia coli]	69	45
238	27	25243	25695	gi 305005	ribulose-5-phosphate 3-epimerase [Spinacia oleracea]	69	52
253	3	1067	921	gi 1591278	ORF_f104 [Escherichia coli]	69	53
260	4	2110	3105	gi 580841	aspartokinase I [Methanococcus jannaschii]	69	39
268	3	2287	1910	gi 460026	F1 [Bacillus subtilis]	69	45
269	7	4532	4083	gi 1303792	repressor protein [Streptococcus pneumoniae]	69	48
					YqeK [Bacillus subtilis]	69	50

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
271	15	11040	12236	gi 1303805	YqeR [Bacillus subtilis]	69	48
271	16	12444	12809	gi 435490	orf1 gene product [Lactococcus lactis]	69	46
281	3	1277	2068	gi 1303968	YqiQ [Bacillus subtilis]	69	50
281	6	5004	5534	gi 1773151	adenine phosphoribosyltransferase [Escherichia coli]	69	54
292	24	19939	18398	gi 1652664	glutamine-binding periplasmic protein [Synechocystis sp.]	69	45
323	3	2708	4243	gi 179401	beta-D-galactosidase precursor (EC 3.2.1.23) [Homo sapiens] gi 179423 beta-galactosidase precursor (EC 3.2.1.23) [Homo sapiens] pir A32688 A32611 beta-galactosidase (EC 3.2.1.23) precursor - uman	69	56
330	2	1388	2353	gi 1303783	YqeC [Bacillus subtilis]	69	48
332	1	2	223	gi 1653594	hemolysin [Synechocystis sp.]	69	50
338	9	7035	7607	gi 467442	stage V sporulation [Bacillus subtilis]	69	55
341	1	1	408	gi 1477741	histidine periplasmic binding protein P29 [Campylobacter jejuni]	69	50
368	2	972	598	gi 516826	rat GCP360 [Rattus rattus]	69	33
375	4	3405	2599	gi 1215693	putative orf; GT9_orf434 [Mycoplasma pneumoniae]	69	38
386	1	2	166	gi 1549376	putative protein [Synechococcus PCC7942]	69	42
396	4	1248	1715	gi 410132	ORFX8 [Bacillus subtilis]	69	50
398	4	2763	2927	gi 466475	putative phospho-beta-glucosidase [Bacillus stearothermophilus] pir D49898 D49898 cellobiose phosphotransferase system celC - acillus stearothermophilus	69	55
421	5	2950	3471	gi 1574625	H. influenzae predicted coding region HI1074 [Haemophilus influenzae]	69	45
423	4	2408	2893	gnl PID e163522	rnhB [Haemophilus influenzae]	69	55
436	3	1763	1521	gi 155032	ORF B [Plasmid pEa34]	69	37

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
452	1	3	341	gi 1591139	M. jannaschii predicted coding region MJ0435 [Methanococcus jannaschii]	69	52
470	3	1816	2181	gi 437389	transposase [Lactococcus lactis]	69	56
471	2	2003	813	gi 854233	cymF gene product [Klebsiella oxytoca]	69	49
478	1	822	4	gi 142521	deoxyribodipyrimidine photolyase [Bacillus subtilis] gnl PID e255102	69	63
					deoxyribodipyrimidine photolyase [Bacillus subtilis]		
490	4	1447	1289	gi 699379	glvr-1 protein [Mycobacterium leprae]	69	41
518	2	213	605	pir S00076 R5BS12	ribosomal protein L12 - Bacillus stearothermophilus	69	59
536	4	1471	1653	gi 1146240	ketopantoate hydroxymethyltransferase [Bacillus subtilis]	69	53
539	5	3796	5091	gi 973231	gamma-glutamyl phosphate reductase [Lycopodium obscurum]	69	54
566	1	1	231	gi 45741	ORFE [Enterococcus faecalis]	69	50
579	5	2729	3595	gi 145887	malonyl coenzyme A-acyl carrier protein transacylase [Escherichia coli]	69	49
583	2	373	912	gi 1064791	function unknown [Bacillus subtilis]	69	55
605	1	254	3	pir S39743 S39743	hypothetical protein - Bacillus subtilis	69	37
630	2	1659	1231	gi 153672	lactose repressor [Streptococcus mutans]	69	47
634	1	36	731	gi 1022725	unknown [Staphylococcus haemolyticus]	69	53
662	1	486	73	gi 467431	high level kasamycin resistance [Bacillus subtilis] sp P37468 KSGA_BACSU	69	55
					DIMETHYLADENOSINE TRANSFERASE (EC 2.1.1.1.-) S-ADENOSYLMETHIONINE-6-N', N'-ADENOSYL(RNA) DIMETHYLTRANSFERASE) 16S RNA DIMETHYLASE) (HIGH LEVEL KASUGAMYCIN RESISTANCE PROTEIN SGA) (K		
689	1	340	26	gi 1017817	membrane spanning protein [Streptomyces coelicolor]	69	41
756	2	300	500	gi 520596	Mre2 protein [Saccharomyces cerevisiae]	69	46



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident
792	2	855	460	gi 1303823	YqfG [Bacillus subtilis]	69	55
916	1	4	789	gnl PID e253114	ornithine carbamoyltransferase [Pyrococcus furiosus]	69	57
7	3	2609	3748	gi 1303836	YqfO [Bacillus subtilis]	68	50
16	5	4165	4689	gi 142450	ahrc protein [Bacillus subtilis]	68	46
17	16	12826	13071	gi 222681	RNA polymerase [Tomato spotted wilt virus]	68	50
17	32	31402	31572	gi 1303984	YqkG [Bacillus subtilis]	68	44
17	33	31509	32009	gi 1303984	YqkG [Bacillus subtilis]	68	50
29	1	19	282	gi 1234787	up-regulated by thyroid hormone in tadpoles; expressed specifically in the tail and only at metamorphosis; membrane bound or extracellular protein; C-terminal basic region [Xenopus laevis]	68	37
29	3	1087	1950	gi 407878	leucine rich protein [Streptococcus equisimilis]	68	45
45	1	204	959	gi 1039479	ORF [Lactococcus lactis]	68	50
47	7	8108	7527	gi 142853	homologous to unidentified E. coli protein [Bacillus subtilis] gi 143161 maf [Bacillus subtilis]	68	46
52	6	4304	5050	gnl PID e124050	alpha-acetolactate decarboxylase [Lactococcus lactis]	68	53
58	5	5961	4807	gi 466365	potential NAD-reducing hydrogenase subunit [Desulfovibrio ructosovorans]	68	49
68	8	4036	4743	gi 1673727	(AE000009) Mycoplasma pneumoniae, glutamine transport ATP-binding protein; similar to Swiss-Prot Accession Number P10346, from E. coli [Mycoplasma pneumoniae]	68	44
72	5	4441	3434	gi 1395209	ribonucleotide reductase R2-2 small subunit [Mycobacterium tuberculosis]	68	52
80	1	836	3	gi 474176	regulator protein [Staphylococcus xylosus]	68	48
81	2	793	1359	gi 1064809	homologous to sp:HTRA_ECOLI [Bacillus]	68	48

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					subtilis]		
85	9	6911	6711	gi 144893	butyrate kinase [Clostridium acetobutylicum]	68	55
89	8	7184	5970	gi 1469784	putative cell division protein ftsw [Enterococcus hirae]	68	44
91	3	828	1076	gi 726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus ubtilis]	68	53
103	1	1019	3	gi 143365	phosphoribosyl aminoimidazole carboxylase II (PUR-K; ttg start odon) [Bacillus subtilis]	68	50
106	2	2441	1509	gi 146860	delta-2-isopentenyl pyrophosphate transferase [Escherichia coli] gi 537012	68	47
					tRNA delta-2-isopentenylpyrophosphate (IPP) transferase [Escherichia coli]		
112	1	558	100	gnl PID e242290	carbamate kinase [Clostridium perfringens]	68	50
116	3	2383	1496	gi 755601	unknown [Bacillus subtilis]	68	42
119	3	2136	1201	gi 1171125	thioredoxin reductase [Clostridium litorale]	68	49
121	4	3697	4650	gi 790945	aryl-alcohol dehydrogenase [Bacillus subtilis]	68	48
123	26	24262	24801	gi 537235	Kenn Rudd identifies as gpmB [Escherichia coli]	68	51
123	27	24887	25888	gi 143150	levR [Bacillus subtilis]	68	51
126	4	2773	1844	gi 551854	ORF2 [Erwinia herbicola]	68	54
131	1	150	1058	gi 1387979	44% identity over 302 residues with hypothetical protein from Synecocystis sp, accession D64006_CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtil	68	44
134	3	2154	1804	sp P39213 YI91_SHI	INSERTION ELEMENT IS911 HYPOTHETICAL 12.7	68	43

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
138	19	12285	12656	DY gi 1438847	KD PROTEIN. homologue of hypothetical 17.6 kDa protein in rplI-cpdB intergenic region of <i>E. coli</i> [Bacillus subtilis]	68	43
151	2	2784	1654	gi 143365	phosphoribosyl aminimidazole carboxylase II (PUR-K; ttg start odon) [Bacillus subtilis]	68	45
164	23	24352	24119	gi 1573564	hypothetical [Haemophilus influenzae]	68	40
166	2	970	1260	gi 151968	nifs [Rhodobacter sphaeroides]	68	41
172	2	1320	2015	gi 1208965	hypothetical 23.3 kd protein [Escherichia coli]	68	46
175	1	900	451	gi 468207	Submitter comments: A Mg2+ transporting P-type ATPase highly homologous with mgfB ATPase at 80 min on Salmonella chromosome. mediates the influx of Mg2+ only. transcription regulated by xtracellular Mg2+ [Salmonella typhimurium]	68	47
180	14	12551	14956	gi 565641	FdRA protein [Escherichia coli]	68	49
186	1	3	686	gi 405804	transposase [Streptococcus thermophilus]	68	51
200	1	239	3	gi 468016	immunoglobulin heavy chain binding protein [Giardia intestinalis]	68	42
201	4	4468	3686	gi 304013	abcA [Aeromonas salmonicida]	68	50
204	10	6833	6468	gi 488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	68	51
214	3	3360	2491	gi 928834	integrase [Lactococcus lactis phage BK5-T]	68	50
229	9	8277	7375	gi 1574569	hypothetical [Haemophilus influenzae]	68	41
229	14	14288	13740	gnl PFD e290287	polypeptide deformylase [Bacillus subtilis]	68	50
230	5	4593	3532	gi 143002	proton glutamate symport protein [Bacillus caldotenax] pir S26246 S26246 glutamate/aspartate transport protein - Bacillus aldotenax	68	29

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
244	1	1	891	gi 537080	ribonucleoside triphosphate reductase [Escherichia coli] pir A47331 A47331 oxygen-sensitive ribonucleoside-triphosphate eductase (EC 1.17.4.-) - Escherichia coli	68	54
244	5	4249	3551	gi 1773172	hypothetical protein [Escherichia coli]	68	46
244	7	5670	5212	gi 467423	unknown [Bacillus subtilis]	68	43
264	9	3925	3734	gi 914991	Similar to hemoglobinase [Saccharomyces cerevisiae] pir S59796 S59796 hypothetical protein D9798.2 - yeast Saccharomyces cerevisiae	68	44
271	7	3484	4686	gi 1469784	putative cell division protein ftsW [Enterococcus hirae]	68	50
271	11	6817	6548	gi 413948	ipa-24d gene product [Bacillus subtilis]	68	50
288	3	1638	1333	gi 562039	NADH dehydrogenase, subunit 2 [Acanthamoeba castellanii] pir S53835 S53835 NADH dehydrogenase chain 2 - Acanthamoeba astellanii mitochondrion (SGC6)	68	50
295	6	3537	4472	gi 555668	glycosylasparaginase precursor [Flavobacterium meningosepticum]	68	41
296	2	3143	1950	gi 1742630	Bicyclomycin resistance protein (Sulfonamide resistance protein). [Escherichia coli]	68	34
301	3	3271	1760	gi 413960	ipa-36d galT gene product [Bacillus subtilis]	68	53
315	3	2230	905	gi 1653498	ABC transporter [Synechocystis sp.]	68	47
318	2	1285	854	gi 43940	EIII-F Sor PTS [Klebsiella pneumoniae]	68	39
320	2	1178	621	gi 664842	sister of P-glycoprotein [Sus scrofa domestica]	68	46
331	2	342	566	pir B48396 B48396	ribosomal protein L33 - Bacillus stearothermophilus	68	59

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
336	1	1	663	gi 1006591	cation-transporting ATPase PacL [Synechocystis sp.]	68	44
338	6	4004	5035	gi 155276	aldehyde dehydrogenase [Vibrio cholerae]	68	51
338	12	10404	11165	gi 467444	transcription-repair coupling factor [Bacillus subtilis] sp P37474 MFD_BACSU. TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF).	68	46
341	3	743	1222	gi 1183886	integral membrane protein [Bacillus subtilis]	68	45
351	6	2992	2561	gi 580881	ipa-73d gene product [Bacillus subtilis]	68	53
363	8	12517	9950	gi 1652980	H(+)-transporting ATPase [Synechocystis sp.]	68	46
368	3	1269	1736	gnl PID e209005	homologous to ORF2 in nrDEF operons of E.coli and S.typhimurium [Lactococcus lactis]	68	37
386	11	6564	6115	gi 765072	ORF3 [Staphylococcus aureus]	68	46
395	3	935	729	gi 15521	ORF 3 (AA 1-90) [Bacteriophage phi-105]	68	34
399	8	6073	6519	gi 153584	biotin carboxyl carrier protein [Streptococcus mutans] sp P29337 BCCP_STRMU BIOTIN CARBOXYL CARRIER PROTEIN (BCCP).	68	53
408	3	2289	1336	gi 41572	GlnP (AA 1-219) [Escherichia coli]	68	40
420	1	559	2	gi 1592142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	68	51
423	2	254	1294	gi 1773109	similar to S. typhimurium apba [Escherichia coli]	68	47
423	3	1465	2421	gi 1653032	hypothetical protein [Synechocystis sp.]	68	40
428	1	859	2	gi 1652454	hypothetical protein [Synechocystis sp.]	68	48
432	7	4626	3901	gi 1573285	hypothetical [Haemophilus influenzae]	68	55
434	1	90	1889	gi 1542975	AbcB [Thermoanaerobacterium thermosulfurigenes]	68	50
441	5	4674	5156	gi 467437	unknown [Bacillus subtilis]	68	48

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
455	4	3835	4080	gi 19815	luminal binding protein (BiP) [Nicotiana tabacum]	68	40
530	2	394	546	gi 763326	unknown [Saccharomyces cerevisiae]	68	42
531	2	810	622	gi 1146183	putative [Bacillus subtilis]	68	51
537	3	1353	1192	gi 929968	ORF; similar to B. anthracis WeyAR element ORF; putative ransposase [Bacillus anthracis]	68	56
539	3	2725	2231	gi 1353537	dUTPase [Bacteriophage r1t]	68	53
569	1	3	446	gi 146544	18 kD protein [Escherichia coli]	68	47
591	2	656	174	gi 1039479	ORFU [Lactococcus lactis]	68	42
652	2	739	1032	gi 1303715	YrkP [Bacillus subtilis]	68	50
671	2	436	1617	gi 413959	ipa-35d galK gene product [Bacillus subtilis]	68	50
684	1	466	2	gnl PID e248400	orfRM1 gene product [Bacillus subtilis]	68	40
693	1	2	787	gi 405804	transposase [Streptococcus thermophilus]	68	46
700	2	772	596	gi 153801	enzyme scr-II [Streptococcus mutans]	68	50
735	1	118	609	gi 969027	gamma-aminobutyrate permease [Bacillus subtilis] sp P46349 GABP_BACSU GABA PERMEASE (4-AMINO BUTYRATE TRANSPORT ARRIER) (GAMA-AMINO BUTYRATE PERMEASE).	68	40
750	1	2	529	gi 893358	PgsA [Bacillus subtilis]	68	54
762	2	1588	950	gi 1146240	ketopantoate hydroxymethyltransferase [Bacillus subtilis]	68	49
790	1	407	3	gi 142224	attachment protein ChvA (ttg strart codon) [Agrobacterium umefaciens]	68	55
882	1	3	278	gi 57572	glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) attus rattus]	68	48
950	1	140	568	gi 882736	ORF f278 [Escherichia coli]	68	53
969	2	554	339	gi 1118031	similar to neural cell adhesion molecules and neuroglins in their IG-like C2-type domains [Caenorhabditis elegans]	68	47

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
970	1	297	73	gi 474404	cyclophilin [Tolypocladium inflatum]	68	40
1	1	1103	3	gi 48790	ORF 3 [Pseudomonas putida]	67	50
29	10	7156	6614	sp P36672 PTTB_ECO LI	PTS SYSTEM, TREHALOSE-SPECIFIC IIBC COMPONENT (EIIBC-TRE) (TREHALOSE- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT) (EC 2.7.1.69) (EII-TRE).	67	52
48	8	8035	9141	gi 975627	N-acylamino acid racemase [Amycolatopsis sp.]	67	48
55	12	6621	7439	gi 391610	farnesyl diphosphate synthase [Bacillus stearothermophilus] pir JX0257 JX0257 geranyltransferase (EC 2.5.1.10) - Bacillus tearothermophilus	67	47
57	13	13972	16401	gnl PID e255138	phenylalanyl-tRNA synthetase beta subunit [Bacillus subtilis]	67	47
63	4	1917	2729	gi 1321629	MIP related protein of E. coli [Escherichia coli]	67	47
68	12	8600	8923	gi 793910	surface antigen [Homo sapiens]	67	43
72	7	7138	6740	gnl PID e209005	homologous to ORF2 in nrDEF operons of E.coli and S.typhimurium [Lactococcus lactis]	67	39
72	10	8309	9433	gi 1199515	ferrous iron transport protein B [Escherichia coli]	67	41
85	5	5315	4296	gi 142611	branched chain alpha-keto acid dehydrogenase E1-alpha [Bacillus ubtilis]	67	52
101	5	4149	3100	gi 1109686	ProX [Bacillus subtilis]	67	48
110	4	2335	1292	gi 1066343	mu-crystallin [Homo sapiens]	67	48
114	12	12936	13520	gi 146218	serine hydroxymethyltransferase [Escherichia coli]	67	50
115	5	3137	2010	gi 1256150	YbaR [Bacillus subtilis]	67	47
115	6	3199	2792	gi 1652593	hypothetical protein [Synechocystis sp.]	67	45
123	25	22739	24208	gi 148711	6-aminohexanoate-cyclic-dimer hydrolase [Flavobacterium sp.] gi 488343 6-	67	50

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					aminohexanoate-cyclic-dimer hydrolase [Flavobacterium p.]		
124	6	5139	4267	gi 1016770	prolipoprotein diacylglycerol transferase [Staphylococcus aureus]	67	50
125	2	1306	221	gi 853743	L-alanoyl-D-glutamate peptidase [Bacteriophage A118]	67	50
128	36	29462	28737	gi 142940	ftsA [Bacillus subtilis]	67	46
138	27	17602	18183	gi 1256639	putative [Bacillus subtilis]	67	50
138	31	21578	20097	gi 143245	Na+/H+ antiporter [Bacillus firmus]	67	42
138	33	25165	23249	gi 1498811	M. jannaschii predicted coding region MJ0050 [Methanococcus jannaschii]	67	45
138	36	28690	27362	gnl PID e269549	Unknown [Bacillus subtilis]	67	47
144	4	3271	3717	gi 1753229	PKC [Borrelia burgdorferi]	67	52
145	3	1435	2511	gi 1573615	ATP-binding protein (abc) [Haemophilus influenzae]	67	47
146	5	4657	2804	gi 1045034	beta-galactosidase [Xanthomonas campestris pv. manihotis]	67	51
149	3	1978	1367	gi 806536	membrane protein [Bacillus acidopullulolyticus]	67	51
156	1	3	365	gnl PID e265539	ClpB-homologue [Thermus aquaticus thermophilus]	67	42
158	15	14863	13766	gi 1573487	rbs repressor (rbsR) [Haemophilus influenzae]	67	40
158	17	16483	15959	gi 677850	hypothetical protein [Staphylococcus aureus]	67	51
159	7	6872	6006	gi 1303949	Yqix [Bacillus subtilis]	67	41
159	9	8103	7498	gi 1303950	Yqiv [Bacillus subtilis]	67	41
165	11	9846	9004	gi 606079	ORF_o267 [Escherichia coli]	67	36
169	2	2151	3047	gi 42371	pyruvate formate-lyase activating enzyme (AA 1-246) [Escherichia li]	67	44
179	13	13648	14451	gnl PID e257631	methyltransferase [Lactococcus lactis]	67	45



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
180	28	28656	29801	gi 666005	hypothetical protein [Bacillus subtilis]	67	48
194	6	2774	4231	gi 143245	Na+/H+ antiporter [Bacillus firmus]	67	41
194	10	6472	8259	gi 622991	mannitol transport protein [Bacillus stearothermophilus] sp P50852 PTMB_BACST PTS SYSTEM, MANNITOL-SPECIFIC IIBC COMPONENT EIIBC-MTL (MANNITOL- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE NYME II, BC COMPONENT) (EC 2.7.1.69) (EII-MTL)..	67	50
204	5	1924	3006	gi 1235684	mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]	67	50
214	1	42	1196	gi 606013	CG Site No. 829 [Escherichia coli]	67	36
219	2	524	850	gnl PID e257628	ORF [Lactococcus lactis]	67	42
223	15	13640	14407	gi 496520	orf iota [Streptococcus pyogenes]	67	54
227	3	1011	1892	gi 1070013	protein-dependent [Bacillus subtilis]	67	37
233	12	9340	8339	gi 507880	xanthine dehydrogenase [Gallus gallus]	67	50
238	10	7951	9183	gi 1653948	hypothetical protein [Synechocystis sp.]	67	45
246	3	783	1430	gnl PID e233869	hypothetical protein [Bacillus subtilis]	67	47
256	2	570	1601	gi 709992	hypothetical protein [Bacillus subtilis]	67	36
266	2	1266	835	gi 963038	ArpU [Enterococcus hirae]	67	42
285	1	3	809	gi 40014	pot. ORF 446 (aa 1-446) [Bacillus subtilis]	67	53
288	10	6838	5801	gi 1651806	hypothetical protein [Synechocystis sp.]	67	45
301	10	8822	8562	gi 1303864	YggQ [Bacillus subtilis]	67	43
312	5	2377	2595	gi 709991	hypothetical protein [Bacillus subtilis]	67	52
353	1	3	1472	gi 151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir A44756 A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	67	48
359	2	984	439	gi 1773190	similar to E. coli yhaE [Escherichia coli]	67	45
359	3	2244	982	gi 1001478	hypothetical protein [Synechocystis sp.]	67	30
364	8	8469	7816	gi 496943	ORF [Saccharomyces cerevisiae]	67	50

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
386	12	6625	7833	gnl PID e254644	membrane protein [Streptococcus pneumoniae]	67	36
394	2	497	2635	gnl PID e255093	hypothetical protein [Bacillus subtilis]	67	45
399	6	5410	3971	gi 665994	hypothetical protein [Bacillus subtilis]	67	45
414	1	1	1227	gi 1621027	high affinity potassium transporter [Debaryomyces occidentalis]	67	40
453	2	618	391	gi 537189	ORF_f132 [Escherichia coli]	67	45
458	1	825	226	gnl PID e189917	ORF 28.5 [Escherichia coli]	67	45
460	2	644	1387	gi 1502421	3-ketoacyl-acyl carrier protein reductase [Bacillus subtilis]	67	48
460	4	2622	3131	gi 1399830	biotin carboxyl carrier protein [Synecococcus PCC7942]	67	53
474	1	1456	77	gi 495277	histidine kinase [Streptococcus pneumoniae]	67	54
488	6	3892	3032	gi 437389	transposase [Lactococcus lactis]	67	47
490	1	460	2	gi 1742830	ORF_ID:0326#2; similar to [SwissProt Accession Number P37794] [Escherichia coli]	67	43
582	1	2	787	gi 1408485	YxdM gene product [Bacillus subtilis]	67	38
629	2	1280	915	gi 1006620	ABC transporter [Synecocystis sp.]	67	50
633	2	941	390	gnl PID e21400	tex gene product [Bordetella pertussis]	67	54
655	1	47	313	gi 147403	mannose permease subunit II-P-Man [Escherichia coli]	67	48
671	3	1630	2415	sp P13226 GALE_STR LI	UDP-GLUCOSE 4-EPIMERASE (EC 5.1.3.2) (GALACTOWALDENASE).	67	52
682	2	1428	595	gi 147404	mannose permease subunit II-M-Man [Escherichia coli]	67	42
704	3	977	411	gi 467428	unknown [Bacillus subtilis]	67	45
711	1	590	168	gi 471236	orf3 [Haemophilus influenzae]	67	37
784	1	253	2	gnl PID e236287	site-specific DNA-methyltransferase [Bacillus stearothermophilus]	67	44

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
907	1	209	3	gi 5119	topoisomerase I [Schizosaccharomyces pombe]	67	42
908	1	275	96	gi 1591045	hypothetical protein (SP:P31466) [Methanococcus jannaschii]	67	46
960	1	499	98	gi 405804	transposase [Streptococcus thermophilus]	67	50
963	1	259	2	pir S34632 S34632	dnaJ protein homolog - human	67	54
964	1	164	628	bbs 173803	CD4+ T cell-stimulating antigen [Listeria monocytogenes, 85EO-1167, Peptide Partial, 268 aa] [Listeria monocytogenes]	67	49
5	4	1438	2403	gi 1303810	YgeT [Bacillus subtilis]	66	50
7	1	24	1727	gi 145220	alanyl-tRNA synthetase [Escherichia coli]	66	50
7	2	1858	2646	gi 687599	orfA1; transposon insertion into orfA1 impairs growth and virulence f L. monocytogenes [Listeria monocytogenes]	66	58
8	1	3	707	gi 1303830	YqfL [Bacillus subtilis]	66	45
9	1	182	1051	gi 467399	IMP dehydrogenase [Bacillus subtilis]	66	51
17	11	8383	8598	gi 457336	Pv200 [Plasmodium vivax]	66	42
18	14	5903	6136	gi 294706	trfA [Plasmid RK2]	66	50
23	12	5951	6895	gi 1652472	ethylene response sensor protein [Synecocystis sp.]	66	51
23	17	11198	11881	gi 466517	pduB [Salmonella typhimurium]	66	44
23	19	12395	13501	gi 145206	alcohol dehydrogenase (adhE) [Escherichia coli]	66	47
34	5	5987	6232	gi 397360	yNucR endo-exonuclease [Saccharomyces cerevisiae]	66	46
43	2	782	1018	gi 513417	non-structural polyprotein of pSP6-SFV4 [unidentified]	66	46
43	5	3757	2324	gnl PID ei54145	penicillin binding protein 4 [Staphylococcus aureus]	66	44
56	4	2351	1662	gi 49272	Asparaginase [Bacillus licheniformis]	66	44
57	2	950	1735	gi 1657505	hypothetical protein [Escherichia coli]	66	46

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
57	4	3117	3932	gi 1657507	hypothetical protein [Escherichia coli]	66	41
57	8	12269	12646	gi 1622733	orf108; unknown function [Butyrivibrio fibrisolvens]	66	44
62	2	547	1302	gi 413967	ipa-43d gene product [Bacillus subtilis]	66	50
62	5	2633	1905	gi 475110	fructokinase [Pediococcus pentosaceus]	66	51
74	7	4661	4086	gi 467484	unknown [Bacillus subtilis]	66	47
81	18	13878	13717	gi 146724	enzyme IIR-Man function protein (manX (ptsL)) [Escherichia coli] gi 41976 manX gene product (AA 1-315) [Escherichia coli]	66	35
94	17	20780	21253	gi 142955	glucose dehydrogenase (EC 1.1.1.47) [Bacillus subtilis] pir S36090 S36090 glucose 1-dehydrogenase (EC 1.1.1.47) - Bacillus ubtilis	66	47
98	15	15165	14338	gi 147327	transport protein [Escherichia coli]	66	34
105	3	1726	3183	gnl PID e205173	orf1 gene product [Lactobacillus helveticus]	66	45
110	17	15811	14804	gi 887824	ORF_o310 [Escherichia coli]	66	52
112	2	712	443	gnl PID e242290	carbamate kinase [Clostridium perfringens]	66	51
123	1	1	540	gi 1573538	H. influenzae predicted coding region HI0552 [Haemophilus influenzae]	66	39
123	33	30312	31460	gi 1498930	M. jannaschii predicted coding region MJ0158 [Methanococcus jannaschii]	66	48
125	8	4914	4474	gi 1736749	Exopolysaccharide production protein PSS. [Escherichia coli]	66	54
128	25	18201	18878	gnl PID e255543	putative iron dependant repressor [Staphylococcus epidermidis]	66	48
131	3	2311	3213	gi 38969	lacF gene product [Agrobacterium radiobacter]	66	37
131	5	3588	3394	gi 1303823	YqfG [Bacillus subtilis]	66	29
135	1	1214	45	gi 1498930	M. jannaschii predicted coding region MJ0158 [Methanococcus jannaschii]	66	48
135	10	7764	7405	gi 530825	OVT1 [Onchocerca volvulus]	66	47

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
144	13	12859	10739	pir A40614 A40614	penicillin-binding protein pbpF - <i>Bacillus subtilis</i>	66	47
145	5	3224	4063	gi 349531	lipoprotein [Pasteurella haemolytica]	66	45
146	2	1497	619	gi 147404	mannose permease subunit II-M-Man [Escherichia coli]	66	38
149	2	1097	1282	gi 1762962	FemA [Staphylococcus simulans]	66	38
150	3	1443	2417	gnl PID e185374	ceuE gene product [Campylobacter coli]	66	46
150	8	6487	6903	gi 1377842	unknown [Bacillus subtilis]	66	43
164	20	21846	22646	gi 1279769	FdhC [Methanobacterium thermoformicicum]	66	57
164	25	24555	25688	pir A43577 A43577	regulatory protein pfor - Clostridium perfringens	66	47
178	1	383	3	gi 763052	integrase [Bacteriophage T270]	66	47
195	19	8698	8516	bbs 169008	homeobox gene [Drosophila sp.]	66	55
207	1	166	1554	gi 619724	MgtE [Bacillus firmus]	66	39
207	3	2312	2010	gi 1204258	soluble protein [Escherichia coli]	66	44
211	3	1523	1729	gi 289932	MHC class II beta chain [Cyphotilapia frontosa]	66	66
213	3	1811	2308	gi 153045	prolipoprotein signal peptidase [Staphylococcus aureus] pir S20433 S20433	66	40
221	7	2524	3468	gi 1353527	ORF10 [Bacteriophage rlt]	66	44
222	13	8272	8988	gi 466719	No definition line found [Escherichia coli]	66	48
223	18	15210	15971	gi 496520	orf iota [Streptococcus pyogenes]	66	57
232	5	3494	2715	gi 142706	comG1 gene product [Bacillus subtilis]	66	41
235	3	1774	734	gi 580897	OppB gene product [Bacillus subtilis]	66	47
244	2	906	1520	gi 15354	ORF 55.9 [Bacteriophage T4]	66	46

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
259	3	2355	1867	gi 56312	Gephyrin [Rattus norvegicus]	66	55
271	1	1	675	gi 1574748	tRNA pseudouridine 55 synthase (truB) [Haemophilus influenzae]	66	53
277	1	1	927	gi 1303799	YqeN [Bacillus subtilis]	66	45
291	5	4587	3547	gnl PID e257609	sugar-binding transport protein [Anaerocellum thermophilum]	66	46
292	25	20451	19912	gi 1649035	high-affinity periplasmic glutamine binding protein [Salmonella typhimurium]	66	50
300	1	2302	77	gi 289262	comE ORF3 [Bacillus subtilis]	66	46
301	4	4290	3265	sp P13226 GALE_STR LI	UDP-GLUCOSE 4-EPIMERASE (EC 5.1.3.2) (GALACTOWALDENASE)	66	51
301	5	4516	4689	gnl PID e212164	PSII, protein N [Odontella sinensis]	66	58
314	1	360	4	gi 467452	unknown [Bacillus subtilis]	66	43
315	4	2559	2209	gi 1653498	ABC transporter [Synechocystis sp.]	66	44
320	3	2406	1081	gnl PID e250352	unknown [Mycobacterium tuberculosis]	66	35
332	2	157	921	gi 1303875	YqhB [Bacillus subtilis]	66	44
334	2	1001	3076	gi 1651660	DNA ligase [Synechocystis sp.]	66	48
338	1	2	616	gi 845686	ORF-27 [Staphylococcus aureus]	66	54
338	7	5011	5496	gi 912476	No definition line found [Escherichia coli]	66	48
341	5	1935	3107	gi 142538	aspartate aminotransferase [Bacillus sp.]	66	44
343	3	2548	2045	gnl PID e289147	similar to single strand binding protein [Bacillus subtilis]	66	44
345	20	22093	22461	gi 1657795	dihydroneopterin aldolase [Methylobacterium extorquens]	66	45
353	3	2621	2379	gnl PID e257628	ORF [Lactococcus lactis]	66	52
365	4	5117	4779	gi 1742868	Mutator MutR protein (7,8-dihydro-8-oxoguanine-triphosphatase) (8-oxo-dgtpase) (EC 3.6.1.-) (DGTP pyrophosphohydrolase). [Escherichia coli]	66	54
376	1	3	1076	gi 1778517	glycerol dehydrogenase homolog	66	45

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					[ <i>Escherichia coli</i> ]		
394	7	5980	5648	gi 486358	ORF_YKL202w [ <i>Saccharomyces cerevisiae</i> ]	66	38
421	4	1469	2539	gi 606375	ORF_f345 [ <i>Escherichia coli</i> ]	66	48
475	6	3978	3763	gi 532547	ORF14 [ <i>Enterococcus faecalis</i> ]	66	48
491	8	7710	7081	gi 1000453	TreR [ <i>Bacillus subtilis</i> ]	66	49
526	1	392	3	gi 1750125	xylose kinase [ <i>Bacillus subtilis</i> ]	66	49
552	6	6147	5917	gi 1432152	PTS antiterminator [ <i>Klebsiella oxytoca</i> ]	66	37
571	2	560	1153	gi 1773132	multidrug resistance-like ATP-binding protein Mdl [ <i>Escherichia coli</i> ]	66	38
575	3	1075	539	gi 1651722	guanylate kinase [ <i>Synechocystis</i> sp.]	66	48
608	2	631	113	gi 1213334	Orfx; hypothetical 22.5 KD protein downstream of type IV prepilin leader peptidase gene; Method: conceptual translation supplied by author [ <i>Vibrio vulnificus</i> ]	66	41
640	1	877	2	sp P50487 YCPX_CLO PE	HYPOTHETICAL PROTEIN IN CPE 5' REGION (FRAGMENT).	66	36
734	1	2	343	gi 1653602	hypothetical protein [ <i>Synechocystis</i> sp.]	66	43
802	1	2	292	gnl PID e280516	voltage-gated sodium channel [Mus musculus]	66	58
812	2	343	531	gi 511075	ORF2 [ <i>Streptococcus agalactiae</i> ]	66	51
823	1	1	393	gi 1303843	YgfV [ <i>Bacillus subtilis</i> ]	66	42
891	1	82	402	gi 567769	ORF5; predicted protein shows similarity to ATP-binding transport proteins AmiE and AmiF of <i>Streptococcus pneumoniae</i> ; disruption of RF5 leads to aminopterin resistance [ <i>Streptococcus parasanguis</i> ]	66	52
5	6	2630	3154	gi 1303811	YgeU [ <i>Bacillus subtilis</i> ]	65	50
6	1	2	628	gi 1742303	Acyl carrier protein phosphodiesterase (ACP phosphodiesterase) (fragment). [ <i>Escherichia coli</i> ]	65	43
18	6	3360	2518	gi 601880	rep protein [ <i>Bacillus borstelensis</i> ]	65	40

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
21	11	7933	7706	gi 1500521	M. jannaschii predicted coding region MJ1623 [Methanococcus jannaschii]	65	32
23	20	13459	13881	gi 488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	65	43
23	25	15987	16178	gnl PID e248966	P32D8.5 [Caenorhabditis elegans]	65	50
27	2	526	302	gi 1001644	regulatory components of sensory transduction system [Synechocystis sp.]	65	44
29	9	6770	5727	sp P36672 PTTB_ECO LI	PTS SYSTEM, TREHALOSE-SPECIFIC IIBC COMPONENT (EIIBC-TRE) (TREHALOSE- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT) (EC 2.7.1.69) (EII-TRE).	65	45
31	5	4611	5207	gi 171625	guanylate kinase [Saccharomyces cerevisiae]	65	39
32	7	4085	3915	gi 150158	29 kD protein [Mycoplasma genitalium]	65	51
33	8	7396	7638	gi 1573421	protein translocation protein, low temperature (secG) [Haemophilus influenzae]	65	26
35	1	2	499	gi 1737500	transcription antiterminator [Bacillus stearothermophilus]	65	40
45	6	2537	3037	gi 511455	unknown [Coxiella burnetii]	65	37
46	3	1028	2254	gi 1001642	dGTP triphosphohydrolase [Synechocystis sp.]	65	43
47	12	14524	14264	gi 150209	ORF 1 [Mycoplasma mycoides]	65	34
50	3	2866	2051	gi 1303830	YqfL [Bacillus subtilis]	65	40
57	11	12955	13332	gnl PID e254999	phenylalanyl-tRNA synthetase beta subunit [Bacillus subtilis]	65	51
62	1	2	484	gi 1573470	H. influenzae predicted coding region HI0491 [Haemophilus influenzae]	65	57
68	1	49	282	gi 1573250	aspartate aminotransferase (aspC) [Haemophilus influenzae]	65	52
72	2	567	1325	gi 466645	alternate name yhiD [Escherichia coli]	65	40
81	5	3711	2938	gi 1732200	PTS permease for mannose subunit IIPMan	65	43



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
83	18	12506	12745	pir D64042 D64042	[Vibrio furnissii] ribosomal-protein-alanine acetyltransferase (rimI) homolog - Haemophilus influenzae (strain Rd KW20)	65	50
100	38	28229	28032	gi 183075	glial fibrillary acidic protein [Homo sapiens]	65	43
105	1	912	106	pir S15248 YQBZCD	fimC protein - Dichelobacter nodosus (serotype D)	65	46
106	5	6097	5102	gi 1143204	ORF2; Method: conceptual translation supplied by author [Shigella sonnei]	65	44
109	3	1165	899	gi 1573390	hypothetical [Haemophilus influenzae]	65	55
110	7	5579	4257	pir B44514 B44514	hypothetical protein 1 (vnfa 5' region) - Azotobacter vinelandii	65	43
120	3	1249	1632	sp P54746 YBGB_ECO LI	HYPOTHETICAL PROTEIN IN HRSA 3'REGION (FRAGMENT)	65	48
122	2	896	1654	gi 1335913	unknown [Erysipelothrix rhusiopathiae]	65	48
145	4	2509	3210	gi 1208965	hypothetical 23.3 kd protein [Escherichia coli]	65	40
149	7	4407	3502	gi 145173	35 kDa protein [Escherichia coli]	65	46
154	8	5738	4926	gi 405804	transposase [Streptococcus thermophilus]	65	47
155	1	306	512	gi 285627	E.coli SecE homologous protein [Bacillus subtilis] pir S39858 S39858 secE protein homolog - Bacillus subtilis sp Q06799 SECE_BACSU PREPROTEIN TRANSLOCASE SECE SUBUNIT.	65	48
158	1	150	1103	gi 289272	ferrichrome-binding protein [Bacillus subtilis]	65	40
158	16	14885	15946	gi 467172	add; L308_C2_206 [Mycobacterium leprae]	65	36
173	4	2103	2912	gnl PID e254877	unknown [Mycobacterium tuberculosis]	65	41
173	12	9749	9054	gi 1652864	hypothetical protein [Synecocystis sp.]	65	50
179	16	15674	17035	gi 1171125	thioredoxin reductase [Clostridium litorale]	65	41

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
180	26	26911	28266	sp P13692 P54_ENTF C	P54 PROTEIN PRECURSOR.	65	39
193	6	2893	3795	gi 39787	adaA [Bacillus subtilis]	65	45
194	5	1843	2238	gi 47394	5-oxopropyl-peptidase [Streptococcus pyogenes]	65	48
199	1	894	82	gi 1591118	nitrate transport ATP-binding protein [Methanococcus jannaschii]	65	46
200	24	13441	13136	gi 144926	toxin A [Clostridium difficile]	65	39
202	3	2925	1846	gi 413968	ipa-44d gene product [Bacillus subtilis]	65	46
203	1	797	3	gi 1377832	unknown [Bacillus subtilis]	65	45
204	3	1065	1472	gi 1008996	unknown [Schizosaccharomyces pombe]	65	51
205	4	1029	1685	gi 148989	truncated tetracycline resistance repressor (non-functional) Haemophilus parainfluenzae]	65	42
206	8	5037	4807	pir D60110 D60110	repetitive protein antigen 3 - Trypanosoma cruzi (fragment)	65	41
217	1	411	4	gi 1146181	putative [Bacillus subtilis]	65	43
217	4	1092	3065	gi 984229	penicillin-binding protein 1a [Streptococcus pneumoniae]	65	48
223	27	23445	23879	gnl PID e269486	Unknown [Bacillus subtilis]	65	47
225	6	5138	3984	gi 39956	ITGlc [Bacillus subtilis]	65	47
229	5	5528	5130	gi 1303914	YqH Y [Bacillus subtilis]	65	33
229	10	10697	8517	gnl PID e266933	unknown [Mycobacterium tuberculosis]	65	46
233	3	2413	1526	gi 887825	ORF_f541 [Escherichia coli]	65	46
236	4	6975	4789	gi 405863	yohA [Escherichia coli]	65	43
237	4	1460	1816	gi 305080	myosin heavy chain [Entamoeba histolytica]	65	42
238	24	21690	23228	gi 305008	rhannulokinase [Escherichia coli]	65	49
242	3	2192	3280	gnl PID e221269	tail protein [Bacteriophage CP-1]	65	37
244	6	5172	4228	gi 1653197	hypothetical protein [Synecocystis sp.]	65	51
259	5	3684	2779	gi 559900	F49E2.1 [Caenorhabditis elegans]	65	39
259	6	4243	3749	gi 1743887	molybdopterin cofactor biosynthesis enzyme	65	50

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident.
260	1	140	478	gi 895748	[Bradyrhizobium japonicum] putative cellobiose phosphotransferase enzyme II' [Bacillus ubtilis]	65	55
269	6	4113	3907	gi 1303792	YqeK [Bacillus subtilis]	65	39
271	12	7731	6772	gi 1657534	cyn operon transcriptional activator [Escherichia coli]	65	45
275	9	6413	5361	gi 1773132	multidrug resistance-like ATP-binding protein Mdl [Escherichia coli]	65	48
276	4	1813	1583	gi 1504014	similar to myosin heavy chain: Containing ATP/GTP-binding site motif A(P-loop) [Homo sapiens]	65	34
279	14	14254	10625	gi 1237015	ORF4 [Bacillus subtilis]	65	45
281	2	692	1279	gi 1303962	YqjK [Bacillus subtilis]	65	50
295	5	2279	3388	gi 436965	[malA] gene products [Bacillus stearothermophilus] pir S43914 S43914 hypothetical protein 1 - Bacillus stearothermophilus	65	41
298	1	63	1142	gi 928834	integrase [Lactococcus lactis phage BK5-T]	65	44
301	8	7592	7176	gi 1303893	YqhL [Bacillus subtilis]	65	50
311	3	4658	5701	gnl PID e221269	tail protein [Bacteriophage CP-1]	65	40
326	1	2	247	gi 466520	pocR [Salmonella typhimurium]	65	38
329	1	789	523	gi 1303895	YqhN [Bacillus subtilis]	65	36
345	5	3363	3641	gi 895749	putative cellobiose phosphotransferase enzyme II' [Bacillus ubtilis]	65	51
369	3	1635	1207	gi 1480429	putative transcriptional regulator [Bacillus stearothermophilus]	65	45
373	2	815	1630	gi 1277032	unknown [Bacillus subtilis]	65	41
379	9	11301	8275	gi 887828	was o492p and o826p before splice [Escherichia coli]	65	49
386	13	7903	8145	gnl PID e217382	M7.9 [Caenorhabditis elegans]	65	39
395	4	1028	1231	gi 1592033	M. jannaschii predicted coding region	65	30

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
396	3	1000	1272	gi 1045900	MJ1387 [Methanococcus jannaschii] hypothetical protein (GB:L09228_17) [Mycoplasma genitalium]	65	44
422	3	2050	1262	gi 405907	yejD [Escherichia coli]	65	50
438	1	44	358	gi 530798	LysB [Bacteriophage phi-LC3]	65	39
460	1	119	646	gi 1502420	malonyl-CoA:Acyl carrier protein transacylase [Bacillus subtilis]	65	46
463	1	870	121	gi 1651917	tRNA(mI37)methyltransferase [Synechocystis sp.]	65	47
468	1	2	823	gi 216457	ORF [Escherichia coli]	65	46
470	1	34	816	gi 530798	LysB [Bacteriophage phi-LC3]	65	47
476	1	21	830	gi 1006591	cation-transporting ATPase PacL [Synechocystis sp.]	65	46
510	7	4875	6092	gi 143150	levR [Bacillus subtilis]	65	46
565	2	686	339	gi 143833	PBSX repressor [Bacillus subtilis]	65	51
566	2	198	743	gi 496501	RepS [Streptococcus pyogenes]	65	34
604	5	1875	2078	gi 1590997	M. jannaschii predicted coding region MJ0272 [Methanococcus jannaschii]	65	49
608	1	194	3	gnl PID e290940	unknown [Mycobacterium tuberculosis]	65	35
648	1	60	953	gi 1591145	hypothetical protein (HI0902) [Methanococcus jannaschii]	65	31
657	4	2531	1620	gi 1500015	amidase [Methanococcus jannaschii]	65	46
691	1	2	718	gnl PID e248400	orfRM1 gene product [Bacillus subtilis]	65	48
704	2	474	175	gi 467428	unknown [Bacillus subtilis]	65	50
758	2	408	683	gi 451201	ORF1 [Bacillus subtilis]	65	44
778	1	833	3	gi 410137	ORFX13 [Bacillus subtilis]	65	40
793	1	1	564	gi 912436	oligo-1,6-glucosidase [Bacillus thermoglucoisidiasius] pir A41707 A41707 oligo-1,6-glucosidase (EC 3.2.1.10) - Bacillus hermoglucoisidiasius	65	40
827	1	364	2	gi 852076	MrgA [Bacillus subtilis]	65	33

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
856	1	209	3	gi 1575605	4-methyl-5-nitrocatechol oxygenase [Burkholderia sp.]	65	45
890	1	966	745	pir A44803 A44803	pG1 protein - human (fragment)	65	63
4	1	2	958	gnl PID e265530	yofE [Streptococcus pneumoniae]	64	43
5	8	4212	5579	gi 407881	stringent response-like protein [Streptococcus equisimilis] pir S39975 S39975 stringent response-like protein - Streptococcus quisiimilis	64	47
8	4	4047	3304	gi 1573150	dihydrolipoamide acetyltransferase (acoC) [Haemophilus influenzae]	64	37
17	14	11709	10393	gi 155109	ORF 1B [Thermus aquaticus thermophilus]	64	37
19	12	6499	6801	gi 1303755	Yqbo [Bacillus subtilis]	64	32
23	1	1	303	gi 1022963	dextranucrase [Leuconostoc mesenteroides]	64	50
28	4	7059	6505	gi 1568609	18kDa protein [Streptococcus pneumoniae]	64	45
31	3	1316	2986	gi 1100076	PTS-dependent enzyme II [Clostridium longisporum]	64	47
47	2	2665	3408	gi 1742154	Phosphoglycolate phosphatase (EC 3.1.3.18). [Escherichia coli]	64	52
48	2	1699	1310	gi 142702	A competence protein 2 [Bacillus subtilis]	64	41
54	8	2750	2352	gi 951052	ORF9, putative [Streptococcus pneumoniae]	64	31
57	15	18035	17274	gi 1183886	integral membrane protein [Bacillus subtilis]	64	40
62	4	1968	1699	gi 475110	fructokinase [Pediococcus pentosaceus]	64	52
100	42	29329	29039	gi 951048	excisionase [Streptococcus pneumoniae]	64	37
102	4	3726	4805	gi 215331	morphogenesis protein [Bacteriophage phi-29]	64	43
106	3	3296	2439	gi 1303930	Yqk [Bacillus subtilis]	64	44
123	12	12960	11314	sp P37047 YAE8_ECO LI	HYPOTHETICAL 44.3 KD PROTEIN IN HTRA-DAPD INTERGENIC REGION.	64	40
128	2	1285	1614	gi 143961	pyruvate phosphate dikinase [Clostridium symbiosum] pir A36231 KIQAPO	64	52

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
128	8	6178	4757	gi 40665	pyruvate,orthophosphate dikinase (EC 2.7.9.1) - lostridium symbiosum		
133	2	1748	2248	gi 1591027	beta-glucosidase [Clostridium thermocellum]	64	41
150	1	35	673	gnl PID e185372	ferripyochelin binding protein [Methanococcus jannaschii]	64	46
158	6	6038	5040	gi 1045801	ceuC gene product [Campylobacter coli]	64	38
164	7	3620	4903	gnl PID e283116	hypothetical protein (SP:P32720) [Mycoplasma genitalium]	64	35
171	11	10107	10784	gi 1591668	unknown similar to quinolon resistance protein NorA [Bacillus subtilis]	64	41
179	4	4826	6373	gi 149535	phosphate transport system regulatory protein [Methanococcus jannaschii]	64	40
181	4	2251	1364	gi 671632	D-alanine activating enzyme [Lactobacillus casei]	64	51
190	11	11302	10355	gi 599850	unknown [Staphylococcus aureus]	64	38
195	37	15344	16033	gi 1736499	orf1 gene product [Lactobacillus sake]	64	33
199	4	4000	5631	gi 746574	Lysostaphin precursor (EC 3.5.1.-) [Escherichia coli]	64	49
202	1	1	1560	gi 309662	similar to M. musculus transport system membrane protein, Nramp PIR:A40739 and S. cerevisiae SMF1 protein (PIR:A45154)	64	37
204	7	3000	4115	gi 1591731	Caenorhabditis elegans]	64	
208	1	308	1090	gi 473821	pheromone binding protein [Plasmid pCF10] jannaschii]	64	45
216	9	6501	6698	gi 47373	melvalonate kinase [Methanococcus jannaschii]	64	41
221	18	8268	8513	gi 1389837	'tetrahydrodipicolinate N-succinyltransferase' [Escherichia coli] gi 1552743 tetrahydrodipicolinate N-succinyltransferase Escherichia coli]	64	42
					7 kDa protein [Streptococcus pneumoniae]	64	35
					complement regulatory protein [Trypanosoma]	64	28

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					Cruzi]		
231	4	2964	2632	gnl PID e279941	muconate cycloisomerase [Rhodococcus erythropolis]	64	37
234	2	751	302	gnl PID e194709	N-terminal part of a protein of unknown function [Chlamydia psittaci]	64	42
238	18	15580	16392	gi 537108	ORF_f254 [Escherichia coli]	64	44
245	1	14	868	gi 153247	endo-beta-N-acetylglucosaminidase H [Streptomyces plicatus] pir A00903 RBSMHP mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) H precursor - treptomyces plicatus	64	51
272	2	584	1144	gi 580781	signal peptidase [Bacillus licheniformis]	64	47
281	5	2659	5019	gi 147550	recJ [Escherichia coli]	64	46
290	12	9496	10371	gi 45713	P.putida genes rpmH, rnpA, 9k, 60k, 50k, gidA, gidB, uncI and uncB pseudomonas putida]	64	42
298	4	4029	3466	gi 147780	rts gene product [Escherichia coli]	64	43
301	20	16216	15977	gi 170482	prosystemin [Solanum lycopersicum]	64	57
301	21	17732	17391	gi 405804	transposase [Streptococcus thermophilus]	64	52
307	1	198	1964	gi 1255196	BSMA [Bacillus stearothermophilus]	64	48
320	5	3441	3070	gi 972900	ArtP [Haemophilus influenzae]	64	38
341	9	7690	6413	gi 1161380	IcaA [Staphylococcus epidermidis]	64	30
345	6	3589	4848	gi 902932	L-methionine gamma-lyase [Pseudomonas putida]	64	45
348	1	453	22	gi 1591957	M. jannaschii predicted coding region MJ1318 [Methanococcus jannaschii]	64	32
350	2	1372	1830	gnl PID e289141	similar to hydroxymyristoyl-(acyl carrier protein) dehydratase [Bacillus subtilis]	64	44
351	7	3291	2917	gi 49013	dTDP-dihydrostreptose synthase [Streptomyces griseus] ir S18618 SYSMPG dTDP-dihydrostreptose synthase - Streptomyces iseus	64	46

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
352	2	780	1028	gi 173431	H+-ATPase [Schizosaccharomyces pombe]	64	38
386	10	5952	6161	gnl PID e243284	ORF YGL056c [Saccharomyces cerevisiae]	64	50
398	2	1233	1808	gi 147920	3-methyladenine-DNA glycosylase I (tag) [Escherichia coli]	64	47
399	12	8761	9159	gi 1778534	HI024 homolog [Escherichia coli]	64	40
409	1	657	1607	gi 1773157	ferrochelatase [Escherichia coli]	64	41
446	1	266	775	gi 563845	orf gene product [Bacillus circulans]	64	53
462	4	1714	1959	gi 169461	serine proteinase inhibitor [Populus trichocarpa x Populus eltooides]	64	50
466	6	5621	8539	gi 143150	levR [Bacillus subtilis]	64	43
501	2	891	1469	gi 467109	rim; 30S Ribosomal protein S18 alanine acetyltransferase; 229_C1_170 [Mycobacterium leprae]	64	44
512	1	1	279	gi 1651948	hypothetical protein [Synechocystis sp.]	64	35
516	1	466	2	gi 155027	6'-N-acetyltransferase [Transposon Tn2426]	64	35
516	2	556	759	gi 1653387	nitrogen assimilation regulatory protein [Synechocystis sp.]	64	58
523	2	904	662	gi 159464	armadillo protein [Musca domestica]	64	45
537	2	1083	844	gi 929966	truncated ORF due to a basepair deletion; similar to B. anthracis terneR element ORF [Bacillus anthracis]	64	42
549	1	309	4	gi 1279769	FdhC [Methanobacterium thermoformicum]	64	48
552	4	5960	3945	gi 1100076	Prs-dependent enzyme II [Clostridium longisporum]	64	47
556	1	3	224	gi 727437	putative 37-kDa protein [Lactococcus lactis]	64	49
557	2	767	1120	gnl PID e257629	transcription factor [Lactococcus lactis]	64	44
602	1	428	156	gi 520407	orf2; GTG start codon [Bacillus thuringiensis]	64	50
603	1	1	165	gi 1621445	sporulation protein Cse15 [Bacillus subtilis]	64	32



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
626	1	3	992	gi 1574715	thioredoxin reductase (trxR) [Haemophilus influenzae]	64	40
628	2	240	446	gi 1165281	Smg [Borrelia burgdorferi]	64	41
723	1	23	829	gi 1620648	surface protein Rib [Streptococcus agalactiae]	64	50
739	1	4	378	gi 143835	PBSX repressor [Bacillus subtilis]	64	37
748	1	139	765	gi 498816	ORF7; homology to regions 4.1 and 4.2 of sigma factors [Bacillus ubtilis]	64	35
758	1	3	410	gi 451201	ORF1 [Bacillus subtilis]	64	34
808	1	368	3	gi 142833	ORF2 [Bacillus subtilis]	64	47
818	2	415	663	gi 854020	U41, major DNA binding protein [Human herpesvirus 6]	64	40
906	1	2	433	gi 1303865	YggR [Bacillus subtilis]	64	44
17	28	28175	27612	gi 151824	ORF5 [Plasmid R46]	63	34
19	18	9546	9722	gi 288661	ORF5 product [Bacteriophage P2]	63	45
39	5	1841	2329	gi 1573292	hypothetical [Haemophilus influenzae]	63	47
41	1	1531	2	gi 580896	nodB protein (aa 1-219) [Bradyrhizobium sp.]	63	43
55	10	5052	6410	gi 1303917	YqiB [Bacillus subtilis]	63	42
80	2	1852	824	gi 38722	precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir A29277 A29277 aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter lcoaceticus	63	42
81	10	6724	6221	gi 1591234	hypothetical protein (SP:P42297) [Methanococcus jannaschii]	63	40
81	14	9175	10848	gi 309662	pheromone binding protein [Plasmid pCF10]	63	44
86	1	2	1006	gi 143316	[gap] gene products [Bacillus megaterium]	63	43
89	13	12929	12639	gi 1377841	unknown [Bacillus subtilis]	63	44
98	14	14365	13502	SP P45169 POTC_HAE IN	SPERMIDINE/PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POTC.	63	37
100	24	20444	17985	gi 563258	virulence-associated protein E	63	44

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					[Dichelobacter nodosus]		
102	2	2441	2599	gi 1619835	MOB [Bacillus thuringiensis israelensis]	63	28
110	22	19725	20705	gi 1763011	lysophospholipase homolog [Homo sapiens]	63	48
115	1	481	92	gi 467360	unknown [Bacillus subtilis]	63	38
128	30	25257	24397	gi 1518679	orf [Bacillus subtilis]	63	39
138	18	12236	11580	gi 405516	This ORF is homologous to nitroreductase from Enterobacter cloacae, ccession Number A38686, and Salmonella, Accession Number P15888 Mycoplasma-like organism]	63	39
143	2	167	1096	pir S39416 S39416	metallothionein 10-I - blue mussel	63	63
158	9	10023	8893	bbs 173803	CD4+ T cell-stimulating antigen [Listeria monocytogenes, 85EO-1167, Peptide Partial, 268 aa] [Listeria monocytogenes]	63	48
164	6	3041	3301	gi 1573583	H. influenzae predicted coding region HI0594 [Haemophilus influenzae]	63	31
164	18	18502	21708	gi 1015903	ORF YJR151c [Saccharomyces cerevisiae]	63	45
165	3	3084	2278	gi 537108	ORF_f254 [Escherichia coli]	63	45
166	1	83	1045	gi 762778	Nifs gene product [Anabaena azollae]	63	49
168	3	638	1489	gi 805022	Ndip [Saccharomyces cerevisiae]	63	32
171	12	10655	10810	gi 152403	phosphate regulatory protein [Rhizobium meliloti]	63	50
172	1	242	1336	gi 1552775	ATP-binding protein [Escherichia coli]	63	45
179	11	11236	12111	gnl PID e245033	unknown [Mycobacterium tuberculosis]	63	42
179	15	15289	15765	gi 1353197	thioredoxin reductase [Eubacterium acidaminophilum]	63	44
180	3	3412	1892	gi 1064813	homologous to sp:PHOR_BACSU [Bacillus subtilis]	63	40
180	7	7063	7926	gi 1657516	hypothetical protein [Escherichia coli]	63	41
187	1	1	729	gi 1651957	hypothetical protein [Synecocystis sp.]	63	34
195	17	7717	8280	gi 431928	MunI methyltransferase [Mycoplasma sp.]	63	44
202	8	5311	6165	gi 606162	ORF_f229 [Escherichia coli]	63	48

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
202	10	7848	8681	gi 606018	ORF_o783 [Escherichia coli]	63	47
208	3	2979	2341	gi 1006613	hypothetical protein [Synechocystis sp.]	63	40
221	3	874	1146	gnl PID e265530	yorFE [Streptococcus pneumoniae]	63	42
227	2	856	1254	gi 438459	homologous to E. coli hydrophobic Fe-uptake components PepD, FecD; utative [Bacillus subtilis]	63	41
231	3	2618	2448	gi 606248	30S ribosomal subunit protein S3 [Escherichia coli]	63	42
233	9	6773	6144	gi 887827	ORF_o192 [Escherichia coli]	63	41
234	1	348	70	gi 494958	Exp2 [Bacillus subtilis]	63	32
240	2	1230	721	gnl PID e252616	DcuC protein [Escherichia coli]	63	38
244	9	7512	6508	gi 467421	similar to B. subtilis DnaH [Bacillus subtilis] sp P37540 YAAS_BACSU HYPOTHETICAL 37.6 KD PROTEIN IN XPAC-ABRB NTERGENIC REGION.	63	43
255	5	3600	2818	gi 1486244	unknown [Bacillus subtilis]	63	47
258	1	3	449	gi 1041115	TRAC [Plasmid pPD1]	63	38
259	4	2842	2342	gnl PID e290788	unknown [Mycobacterium tuberculosis]	63	42
265	8	3313	3480	gi 694074	emml gene product [Streptococcus pyogenes]	63	42
276	18	12505	11654	gi 601878	beta-1,3-glucanase bglH [Bacillus circulans]	63	36
294	5	2012	2275	gi 288661	ORF5 product [Bacteriophage P2]	63	40
301	7	7063	6704	gnl PID e290998	unknown [Mycobacterium tuberculosis]	63	41
345	2	2279	2725	gi 413940	ipa-16d gene product [Bacillus subtilis]	63	39
351	8	4361	3306	gi 398120	TDP-glucose oxidoreductase [Xanthomonas campestris]	63	47
359	1	526	14	gi 1001605	3-hydroxyisobutyrate dehydrogenase [Synechocystis sp.]	63	36
364	6	6741	7277	gi 1736473	ORF_ID:0335#13; similar to [SwissProt Accession Number P36088] [Escherichia coli]	63	42

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
378	2	683	1414	gi 529016	aminoglycoside 6-adenylyltransferase [Bacillus subtilis] pir JU0059 XXBSG aminoglycoside 6-adenylyltransferase (EC 2.7.7.-) Bacillus subtilis	63	41
392	2	783	1646	gi 1772644	orfR gene product [Bacillus subtilis]	63	34
399	2	574	1407	gi 40023	B.subtilis genes rpmH, rnpA, 50kd, gidA and gidB [Bacillus subtilis] i 467388 stage III sporulation [Bacillus subtilis] ir S18073 S18073 spoIIJ protein - Bacillus subtilis	63	42
403	1	754	2	gi 1303938	YgiS [Bacillus subtilis]	63	52
404	5	4149	3745	gi 142450	ahrC protein [Bacillus subtilis]	63	42
430	1	2	1222	gi 1046082	M. genitalium predicted coding region MG372 [Mycoplasma genitalium]	63	40
432	1	3	1241	gi 1001328	UDP-MurNac-tripeptide synthetase [Synechocystis sp.]	63	33
432	4	1970	3016	gi 1161061	dioxygenase [Methylobacterium extorquens]	63	41
463	2	1324	851	gi 1573163	hypothetical [Haemophilus influenzae]	63	40
466	4	2843	3730	gnl PID e261988	putative ORF [Bacillus subtilis]	63	41
472	1	527	3	gi 556885	Unknown [Bacillus subtilis]	63	50
517	3	2803	1646	gi 531265	lipophilic protein which affects bacterial lysis rate and ethcillin resistance level [Staphylococcus aureus] pir A55856 A55856 llm protein - Staphylococcus aureus	63	38
538	1	206	3	gi 172657	serine-protein kinase [Saccharomyces cerevisiae]	63	47
539	4	2997	3851	gi 973230	gamma-glutamyl kinase [Lycopersicon esculentum]	63	43
565	3	756	1010	gi 1303724	YqaF [Bacillus subtilis]	63	51
573	7	4518	3709	gi 1652352	dihydropteroate pyrophosphorylase [Synechocystis sp.]	63	45
579	2	361	1344	gi 1573114	beta-ketoacyl-acyl carrier protein	63	41

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					synthase III (fabH) [Haemophilus influenzae]		
593	2	390	1037	gi 409286	bmrU [Bacillus subtilis]	63	33
707	1	647	171	gi 511596	interleukin-2 [Canis familiaris]	63	33
714	1	2	268	gnl PID e213832	putative inner membrane protein [Bacillus licheniformis]	63	38
724	1	562	239	gnl PID e255315	unknown [Mycobacterium tuberculosis]	63	49
759	1	681	4	gi 437639	[Plasmodium falciparum 3' end.], gene product [Plasmodium alciparum]	63	28
794	1	981	313	gi 451201	ORF1 [Bacillus subtilis]	63	37
811	2	609	184	gi 150553	regulatory protein [Plasmid pCF10]	63	30
835	1	2	262	gi 1736496	RpiR protein. [Escherichia coli]	63	41
11	1	2	1144	gi 143150	levR [Bacillus subtilis]	62	48
12	5	8710	7673	gi 1486244	unknown [Bacillus subtilis]	62	43
15	3	1167	2957	gi 1592101	adenine deaminase [Methanococcus jannaschii]	62	40
16	4	2572	4092	gi 1109685	ProW [Bacillus subtilis]	62	37
23	4	1279	2067	gi 41432	fepC gene product [Escherichia coli]	62	35
23	26	16176	16454	gi 154499	carbon dioxide concentrating mechanism protein [Synechococcus sp.] pir C36904 C36904 carbon dioxide concentrating mechanism protein cml - Synechococcus sp. (PCC 7942)	62	41
31	6	5322	5774	gi 532309	25 kDa protein [Escherichia coli]	62	38
68	4	1606	2778	gi 1732203	GlcNAc 6-P deacetylase [Vibrio furnissii]	62	44
72	1	1	540	gi 1573097	glucosamine-6-phosphate deaminase protein (nagB) [Haemophilus influenzae]	62	26
76	3	1937	2227	gi 928830	ORF75; putative [Lactococcus lactis phage BK5-T]	62	34
83	16	11700	12272	gi 1592161	N-terminal acetyltransferase complex, subunit ARD1 [Methanococcus jannaschii]	62	33

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
83	19	12685	13737	gi 1653193	sialoglycoprotease [Synecocystis sp.]	62	42
91	6	3232	3789	gi 1762962	FemA [Staphylococcus simulans]	62	37
100	43	29676	29317	gi 963033	orf1 gene product [Enterococcus hirae]	62	45
101	8	7410	6481	gi 1161061	dioxygenase [Methylobacterium extorquens]	62	45
110	3	653	871	gi 992683	mdm2-D [Homo sapiens]	62	37
110	8	8440	5810	gi 784897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir A56390 A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	62	46
111	2	1057	287	gnl PID e253280	ORF YDL238c [Saccharomyces cerevisiae]	62	45
114	5	6886	7662	gi 152719	flavocytochrome c [Shewanella putrefaciens]	62	37
115	4	1401	1994	gi 1303978	YqkA [Bacillus subtilis]	62	46
118	1	545	225	gi 39431	oligo-1,6-glucosidase [Bacillus cereus]	62	40
119	8	4625	4356	gi 1522673	type I restriction enzyme [Methanococcus jannaschii]	62	33
120	2	257	1270	gnl PID e235823	unknown [Schizosaccharomyces pombe]	62	41
121	8	7543	8034	gi 39475	formamidopyrimidine-DNA glycosylase [Bacillus firmus] ir S11489 S11489 formamidopyrimidine-DNA glycosidase (EC 3.2.2.23) Bacillus firmus	62	48
123	2	1677	592	gi 882252	conjugated bile acid hydrolase [Clostridium perfringens] sp P54965 CBH_CLOPE CHOLYLGLYCINE HYDROLASE (EC 3.5.1.24) CONJUGATED BILE ACID HYDROLASE (CBAH) (BILE SALT HYDROLASE).	62	40
128	16	10895	9408	gi 1742834	PTS system, cellobiose-specific IIC component (EIIIC-CEL) (Cellobiose- permease IIC component) (Phosphotransferase enzyme II, C component). [Escherichia coli]	62	43

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
128	29	24254	23544	gi 1518680	minicell-associated protein DivIVA [Bacillus subtilis]	62	37
128	35	28843	28103	gi 142940	ftsA [Bacillus subtilis]	62	42
133	4	3434	4165	gnl PID e235174	unknown [Mycobacterium tuberculosis]	62	38
134	2	1679	933	gi 155032	ORF B [Plasmid pEa34]	62	36
146	6	4923	4651	gi 153675	tagatose 6-P kinase [Streptococcus mutans]	62	48
149	5	3318	2527	gi 1591587	pantothenate metabolism flavoprotein [Methanococcus jannaschii]	62	35
152	9	4830	5747	gi 1652461	lactose transport system permease protein LacF [Synechocystis sp.]	62	39
163	2	1341	544	gi 1533098	DnaD protein [Bacillus subtilis]	62	41
164	14	9567	9322	gi 1118060	coded for by C. elegans cDNA yk3dl1.5; coded for by C. elegans cDNA yk5f4.5 [Caenorhabditis elegans]	62	27
172	8	6613	7146	gi 915199	ggaB [Bacillus subtilis]	62	33
173	13	11127	9736	gi 1653484	hypothetical protein [Synechocystis sp.]	62	44
177	1	1077	364	gi 1572994	2-keto-3-deoxy-6-phosphogluconate aldolase (eda) [Haemophilus influenzae]	62	38
178	4	1683	1318	gnl PID e155310	Orf2 [Bacteriophage TP901-1]	62	51
179	5	6425	7576	gi 1161933	DltB [Lactobacillus casei]	62	44
180	13	12470	10842	sp P37047 YAEG_ECO LI	HYPOTHETICAL 44.3 KD PROTEIN IN HTA-DAPD INTERGENIC REGION.	62	38
181	14	11649	10735	gi 1742758	Shikimate 5-dehydrogenase (EC 1.1.1.25). [Escherichia coli]	62	41
197	2	516	1442	gi 623476	transcriptional activator [Providencia stuartii] sp P43463 AARP_PROST TRANSCRIPTIONAL ACTIVATOR AARP.	62	34
206	5	2728	1790	gnl PID e265638	unknown [Mycobacterium tuberculosis]	62	37
210	2	938	2290	gi 528991	unknown [Bacillus subtilis]	62	41
221	15	7083	7280	gnl PID e219154	K08F4.5 [Caenorhabditis elegans]	62	44
222	11	7141	8022	gi 537034	ORF_o488 [Escherichia coli]	62	39

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
223	9	6924	6358	gnl PID e283128	unknown, highly similar to <i>E. coli</i> YecD hypothetical 21.8 KP protein in asps 5'-region and to isochorismatase [ <i>Bacillus subtilis</i> ]	62	42
225	4	2055	2885	gi 18724	pyrroline-5-carboxylate reductase (AA 1-274) [Glycine max] ir S10186 S10186 pyrroline-5-carboxylate reductase (EC 1.5.1.2) - ybean	62	39
229	11	11428	10670	gnl PID e235745	hypothetical protein [ <i>Mycobacterium leprae</i> ]	62	36
231	1	1244	3	gi 48808	dciAE gene product [ <i>Bacillus subtilis</i> ]	62	45
233	1	801	4	gi 143391	ORF2 [ <i>Bacillus subtilis</i> ]	62	42
233	13	10471	9431	gi 887825	ORF_f541 [ <i>Escherichia coli</i> ]	62	35
242	1	3	149	gi 532549	ORF16 [ <i>Enterococcus faecalis</i> ]	62	44
255	2	443	1009	gi 639789	ORF9 [ <i>Mycoplasma pneumoniae</i> ]	62	44
266	6	2349	2158	gnl PID e194945	yeast sds22 homolog [ <i>Homo sapiens</i> ]	62	37
270	1	3	314	gi 1303827	YqfI [ <i>Bacillus subtilis</i> ]	62	35
270	7	5136	4447	gi 1303958	YqjG [ <i>Bacillus subtilis</i> ]	62	41
279	1	271	2	gnl PID e185372	ceuC gene product [ <i>Campylobacter coli</i> ]	62	44
301	11	9598	8798	gi 1303863	YggP [ <i>Bacillus subtilis</i> ]	62	45
306	2	750	1202	gi 148771	ribosomal protein HmaS4 [ <i>Haloarcula marismortui</i> ]	62	41
308	3	2328	1684	gnl PID e238666	hypothetical protein [ <i>Bacillus subtilis</i> ]	62	40
309	5	8806	8573	gi 1591861	M. jannaschii predicted coding region MJ1230 [ <i>Methanococcus jannaschii</i> ]	62	37
318	3	2278	1283	gi 1256134	YbbE [ <i>Bacillus subtilis</i> ]	62	37
321	3	1433	1792	gi 606080	ORF_o290; Geneplot suggests frameshift linking to o267, not found <i>Escherichia coli</i>	62	37
338	13	11175	12770	gi 467446	similar to SpoVB [ <i>Bacillus subtilis</i> ]	62	38
345	11	10519	11793	gi 1736789	Collagenase precursor (EC 3.4.-.-).	62	40



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
345	21	22459	22947	gi 1657794	[Escherichia coli] 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase [Methylobacterium extorquens]	62	47
358	1	902	36	gi 409241	penicillin-binding protein 2 [Staphylococcus aureus]	62	44
362	6	2930	3493	gnl PID e255091	hypothetical protein [Bacillus subtilis]	62	37
363	2	3242	1581	gnl PID e254997	hypothetical protein [Bacillus subtilis]	62	40
365	2	400	1770	gi 143150	levR [Bacillus subtilis]	62	42
372	5	2525	4489	gi 1045736	fructose-permease IIBC component [Mycoplasma genitalium]	62	43
373	1	3	851	gi 438462	transmembrane protein [Bacillus subtilis]	62	36
375	1	2	1336	gi 732813	branched-chain amino acid carrier [Lactobacillus delbrueckii]	62	43
375	3	2592	1831	gi 1644206	unknown [Bacillus subtilis]	62	43
391	2	142	510	gi 151776	ORF3 [Escherichia coli]	62	31
396	2	254	1051	gi 410131	ORFX7 [Bacillus subtilis]	62	41
423	1	197	6	pir A33592 A33592	repressor protein catM - Acinetobacter calcoaceticus	62	38
436	1	704	3	gi 455376	unidentified reading frame L (ORFL) (putative); putative [Transposon n10]	62	32
466	8	9320	10480	gi 147402	mannose permease subunit III-Man [Escherichia coli]	62	44
488	5	2175	2927	gi 532546	ORF13 [Enterococcus faecalis]	62	40
510	4	2572	3078	gi 43941	EIII-B Sor PTS [Klebsiella pneumoniae]	62	35
517	2	1533	736	gi 559388	epsX gene product [Acinetobacter calcoaceticus]	62	53
519	1	2	1084	gi 1652876	hypothetical protein [Synechocystis sp.]	62	41

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
535	1	353	69	gi 1196922	unknown protein [Insertion sequence IS861]	62	33
579	1	1	363	gi 535052	involved in protein secretion [Bacillus subtilis]	62	22
656	5	5351	5956	gnl PID e290931	unknown [Mycobacterium tuberculosis]	62	40
666	1	445	128	gi 483940	transcription regulator [Bacillus subtilis]	62	42
682	1	597	172	gi 146724	enzyme IIR-Man function protein (manX (ptsL)) [Escherichia coli] gi 41976 manX gene product (AA 1-315) [Escherichia coli]	62	37
771	1	3	365	gi 1773086	similar to S. typhimurium ProY [Escherichia coli]	62	44
831	1	390	94	gnl PID e255000	hypothetical protein [Bacillus subtilis]	62	55
15	5	4421	5260	gnl PID e214719	PlcR protein [Bacillus thuringiensis]	61	38
16	6	4705	4938	gi 758425	complement component C3 [Xenopus laevis/gilli]	61	44
23	16	10279	11214	sp P19265 EUTC_SAL TY	ETHANOLAMINE AMMONIA-LYASE LIGHT CHAIN (EC 4.3.1.7).	61	46
33	2	1789	2205	gi 413958	ipa-34d gene product [Bacillus subtilis]	61	36
33	5	4756	6594	gi 1001823	cadmium-transporting ATPase [Synechocystis sp.]	61	38
37	4	2813	3295	gi 1256140	YbbK [Bacillus subtilis]	61	51
37	7	5973	5215	gnl PID e269488	Unknown [Bacillus subtilis]	61	33
49	4	1567	1839	gnl PID e139445	major tail protein [Bacteriophage B1]	61	43
56	1	108	641	gi 1574067	H. influenzae predicted coding region HI1034 [Haemophilus influenzae]	61	35
59	1	1	1002	gi 763513	ORF4; putative [Streptomyces violaceoruber]	61	37
69	7	4837	5523	gnl PID e254877	unknown [Mycobacterium tuberculosis]	61	34
72	11	9262	10476	gi 1591272	ferrous iron transport protein B [Methanococcus jannaschii]	61	45
83	2	731	1549	gi 755152	highly hydrophobic integral membrane	61	41

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					protein [Bacillus subtilis] sp P42953 TAGG_BACSU TEICHOIC ACID TRANSLLOCATION PERMEASE PROTEIN AGG.		
87	2	2067	925	gi 1573129	hypothetical [Haemophilus influenzae]	61	46
103	5	2689	3495	gi 1685111	orf1091 [Streptococcus thermophilus]	61	45
110	13	11455	11820	gi 1001825	transcriptional repressor SmtB [Synechocystis sp.]	61	42
110	15	14048	12588	gi 1573583	H. influenzae predicted coding region HI0594 [Haemophilus influenzae]	61	38
111	3	1675	1055	gnl PID e253280	ORF YDL238c [Saccharomyces cerevisiae]	61	34
111	4	1838	2518	gi 1574513	hypothetical [Haemophilus influenzae]	61	50
111	5	2535	3158	gi 537235	Kenn Rudd identifies as gpMB [Escherichia coli]	61	40
121	1	3	1397	gi 290643	ATPase [Enterococcus hirae]	61	50
123	28	25608	27734	gi 143150	levR [Bacillus subtilis]	61	39
125	5	3455	2589	gi 148921	LicD protein [Haemophilus influenzae]	61	47
128	14	9382	9146	gi 575361	protein kinase PkPA [Phycomyces blakesleeanus]	61	38
138	32	23151	21628	gi 1184262	GadC [Shigella flexneri]	61	34
144	8	6311	5325	gi 710422	cmp-binding-factor 1 [Staphylococcus aureus]	61	39
171	4	4601	5566	gi 41500	ORF 3 (AA 1-352); 38 kD (put. ftsX) [Escherichia coli]	61	31
172	3	2006	2848	gi 303560	ORF271 [Escherichia coli]	61	42
173	7	5146	6228	gi 1256134	YbbE [Bacillus subtilis]	61	31
197	8	9183	8182	gi 143803	GerC3 [Bacillus subtilis]	61	33
217	5	3007	3462	gi 1749414	unnamed protein product [Schizosaccharomyces pombe]	61	43
217	8	6099	5464	gi 143456	rpoE protein (ttg start codon) [Bacillus subtilis]	61	37
222	6	3400	3927	gnl PID e255118	hypothetical protein [Bacillus subtilis]	61	41

Table 3: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
225	3	1946	981	gi 1574660	xylose operon regulatory protein (xylR) [Haemophilus influenzae]	61	43
237	2	203	952	gi 1019108	alternate start at bp 59; ORF [Bacteriophage phi-80]	61	52
237	7	3058	3279	gnl PID e246904	ORF YPL169c [Saccharomyces cerevisiae]	61	32
262	1	20	913	gnl PID e214719	PlcR protein [Bacillus thuringiensis]	61	35
271	17	12725	13504	gi 143057	ORF39 [Bacillus subtilis]	61	31
275	8	5370	3697	gi 1542975	AbcB [Thermoanaerobacterium thermosulfurigenes]	61	41
280	2	692	3079	gi 1001352	ABC transporter [Synechocystis sp.]	61	42
294	7	2276	2767	gi 662792	single-stranded DNA binding protein [unidentified eubacterium]	61	44
301	12	9965	9519	gi 1303861	YggN [Bacillus subtilis]	61	41
308	1	1471	26	gi 1276882	EpsI [Streptococcus thermophilus]	61	36
314	2	475	1662	gi 1975351	PatB [Bacillus subtilis]	61	42
321	9	3762	4193	gi 1732202	PFS permease for mannose subunit IirMan N terminal domain [Vibrio furnissii]	61	40
323	5	5118	5537	gi 532540	ORF7 [Enterococcus faecalis]	61	28
324	7	4800	5156	gi 146122	H-protein [Escherichia coli]	61	39
338	3	1456	1989	pir A47071 A47071	orf1 immediately 5' of rifs - Bacillus subtilis	61	43
341	2	342	947	gi 1736577	Octopine transport system permease protein OccM. [Escherichia coli]	61	41
349	3	1788	1363	pir G64143 G64143	hypothetical protein HI0143 - Haemophilus influenzae (strain Rd KW20)	61	38
369	2	1261	587	gi 153744	ORF X; putative [Streptococcus mutans]	61	33
371	2	1801	1562	gi 48836	xylulokinase [Staphylococcus xylosum]	61	40
372	4	1575	2543	gi 149395	lacC [Lactococcus lactis]	61	43
379	11	12683	11727	gi 887829	D21141 uses 2nd start; frame determined by Lac fusion [Escherichia coli]	61	40
383	5	5625	3820	gi 624072	similar to Escherichia coli	61	36

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
					glycerophosphoryl diester hosphodiesterase, Swiss-Prot Accession Number P10908 [Paramecium ursaria Chlorella virus 1]		
395	2	771	517	gnl PID e276251	T23G11.6 [Caenorhabditis elegans]	61	42
399	20	15621	15812	gi 472527	protein phosphatase 1 [Schizosaccharomyces pombe]	61	44
413	1	3	749	gnl PID e289144	ywpE [Bacillus subtilis]	61	42
427	1	1079	288	gi 403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir S37251 S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	61	42
436	4	2045	1761	gi 48669	pot. ORF B [Shigella sonnei]	61	38
437	1	1158	244	gi 580866	ipa-12d gene product [Bacillus subtilis]	61	47
482	2	1676	1167	bbs 158786	4A11 antigen, sperm tail membrane antigen=putative sucrose-specific phosphotransferase enzyme II homolog [mice, testis, Peptide Partial, 172 aa] [Mus sp.]	61	42
490	3	1291	1094	gnl PID e248473	putative phosphate permease [Arabidopsis thaliana]	61	35
514	1	687	142	gi 1742775	msm operon regulatory protein. [Escherichia coli]	61	36
541	1	758	3	gi 1591732	cobalt transport ATP-binding protein O [Methanococcus jannaschii]	61	39
551	3	2163	1600	gi 671632	unknown [Staphylococcus aureus]	61	38
603	2	163	564	gi 1408587	relaxase [Lactococcus lactis lactis]	61	39
637	8	4539	4769	gi 143559	subtilin [Bacillus subtilis]	61	38
765	1	34	681	gi 408888	orfa 5' of intG [lactobacillus bacteriophage phi adh] pir PN0468 PN0468 hypothetical protein 106 - Lactobacillus	61	40

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					gasserii fragment)		
773	1	53	1207	gi 143841	xylose repressor [Bacillus subtilis]	61	36
798	1	175	381	gi 187572	located at OATL1 [Homo sapiens]	61	32
5	2	303	998	gi 1783264	homologous to DNA glycosylases; hypothetical [Bacillus subtilis]	60	50
8	8	5891	6550	gi 1777939	Pfs [Treponema pallidum]	60	40
11	7	4096	4935	gi 147404	mannose permease subunit II-M-Man [Escherichia coli]	60	41
11	8	4919	5254	gi 467125	glms; L-Glucosamine:D-fructose-6-phosphate aminotransferase; 229_C3_238 [Mycobacterium leprae]	60	30
17	9	7736	8203	gi 496514	orf zeta [Streptococcus pyogenes]	60	42
20	1	3	443	gi 861137	chitin binding protein [Streptomyces olivaceoviridis] pir S55001 S55001 CHB1 protein - Streptomyces olivaceoviridis (SUB -30)	60	40
21	3	1970	684	gi 1778520	hypothetical protein [Escherichia coli]	60	43
23	11	5357	5953	gi 619066	NAST [Azotobacter vinelandii]	60	31
34	4	6662	3279	gi 153952	polymerase III polymerase subunit (dnaE) [Salmonella typhimurium] pir A45915 A45915 DNA-directed DNA polymerase (EC 2.7.7.7) III lpha chain - Salmonella typhimurium	60	37
39	1	47	466	gi 1561567	Unknown [Bacillus subtilis]	60	35
39	4	1855	1361	gi 298045	Orf154 [Streptomyces ambofaciens]	60	41
48	4	2554	4128	gi 1255259	o-succinylbenzoic acid (OSB) CoA ligase [Staphylococcus aureus]	60	40
56	9	6682	5795	gi 413940	ipa-16d gene product [Bacillus subtilis]	60	40
65	3	2105	2593	gi 1573061	hypothetical [Haemophilus influenzae]	60	34
72	9	7854	8330	gi 606343	CG Site No. 28964 [Escherichia coli]	60	39
81	3	2053	1406	gi 1574770	phenylalanyl-tRNA synthetase beta-subunit (pheT) [Haemophilus influenzae]	60	46

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
81	4	2987	2130	gi 147404	mannose permease subunit II-M-Man [Escherichia coli]	60	34
81	12	8280	7150	gnl PID e254984	hypothetical protein [Bacillus subtilis]	60	44
83	22	16887	16537	gi 509672	repressor protein [Bacteriophage Tuc2009]	60	33
89	1	698	60	gi 840838	hypothetical 21.7 kDa protein in ftsY 5' region [Pseudomonas eruginosa]	60	36
89	12	12641	11856	gi 1377843	unknown [Bacillus subtilis]	60	40
89	17	18879	15844	gi 666069	orf2 gene product [Lactobacillus leichmannii]	60	37
94	6	2281	3384	gi 468760	ORF334 [Rhizobium meliloti]	60	36
98	1	12	1970	gi 1652892	ABC transporter [Synechocystis sp.]	60	38
99	3	978	1460	gi 473955	DNA-binding protein [Lactobacillus sp.]	60	31
100	35	26818	26333	gi 347851	junctional sarcoplasmic reticulum glycoprotein [Oryctolagus unicolor]	60	48
100	45	30072	30449	gi 143547	Sin regulatory protein (ttg start codon) [Bacillus subtilis] gi 1303886 SinR [Bacillus subtilis]	60	43
102	8	5923	6561	gi 1633572	Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus]	60	25
109	1	362	3	pir S10655 S10655	hypothetical protein X - Pyrococcus woesei (fragment)	60	33
110	16	14806	14087	pir JH0364 JH0364	hypothetical protein 176 (SAGP 5' region) - Streptococcus pyogenes	60	35
110	20	18929	18414	gi 142450	ahrC protein [Bacillus subtilis]	60	39
110	21	19124	19624	gi 142450	ahrC protein [Bacillus subtilis]	60	40
111	1	289	2	gi 1256618	transport protein [Bacillus subtilis]	60	31
122	7	5627	9589	gi 217191	5'-nucleotidase precursor [Vibrio parahaemolyticus]	60	39
123	5	4390	3659	gi 1197667	vitellogenin [Anolis pulchellus]	60	27
123	20	18102	18407	gi 1303705	YrkF [Bacillus subtilis]	60	34

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
128	32	26229	25492	gi 1652485	hypothetical protein [Synechocystis sp.]	60	29
129	5	4421	6259	gi 1303853	YggF [Bacillus subtilis]	60	36
131	2	1112	2338	gi 699112	ugpC gene product [Mycobacterium leprae]	60	41
131	4	3194	4036	gi 296356	putative membrane transport protein [Clostridium perfringens]	60	32
131	8	6669	7901	gi 537054	pir A56641 A56641 probable membrane transport protein - Clostridium erfringens 2',3'-cyclic-nucleotide 2'-phosphodiesterase [Escherichia coli] pir S56438 S56438.2',3'-cyclic-nucleotide 2'-phosphodiesterase (EC 1.4.16) - Escherichia coli	60	40
133	11	9854	10240	gnl PID e249654	YneR [Bacillus subtilis]	60	37
138	7	6793	6263	gi 1486247	unknown [Bacillus subtilis]	60	48
146	4	2831	2328	gi 39979	P18 [Bacillus subtilis]	60	38
149	6	3504	3316	gi 145173	35 kDa protein [Escherichia coli]	60	47
154	5	2599	3558	gi 1773109	similar to S. typhimurium appa [Escherichia coli]	60	41
155	5	3061	4701	gi 388269	traC [Plasmid pAD1]	60	38
155	11	8565	8927	gi 1197460	MtFB [Escherichia coli]	60	39
158	10	11123	10032	gi 581809	tmbC gene product [Treponema pallidum]	60	39
165	7	6131	5700	gi 1439527	EIIA-man [Lactobacillus curvatus]	60	35
172	4	3169	3810	gi 1001342	hypothetical protein [Synechocystis sp.]	60	42
174	2	1574	762	gi 1045808	hypothetical protein (GB:U00021_19) [Mycoplasma genitalium]	60	35
181	7	4975	4460	gi 683584	shikimate kinase [Lactococcus lactis]	60	33
183	6	2719	2955	gi 1146198	ferredoxin [Bacillus subtilis]	60	37
189	2	3528	2221	gi 396301	matches P500041: Bacterial regulatory proteins, araC family ignature [Escherichia coli]	60	35
193	5	3121	2600	gi 39788	adaB [Bacillus subtilis]	60	49



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
195	11	4623	6569	gnl PID e250887	potential coding region [Clostridium difficile]	60	39
202	2	1837	1607	gi 693939	membrane ATPase [Haloferax volcanii]	60	32
206	7	4794	3754	gi 1574702	hypothetical [Haemophilus influenzae]	60	42
209	2	1308	433	pir A38587 A38587	collagen, corneal - chicken (fragment)	60	51
220	3	4263	1213	gi 437706	alternative truncated translation product from E.coli [Streptococcus pneumoniae]	60	41
222	9	6019	6522	gi 882463	protein-N(pi)-phosphohistidine-sugar phosphotransferase [Escherichia coli]	60	47
222	12	8001	8336	gi 537035	ORF_0101 [Escherichia coli]	60	33
233	2	1294	827	gi 145091	flavodoxin [Desulfovibrio salexigens]	60	39
242	11	7370	7627	gi 1353404	cytochrome oxidase subunit I [Metridium senile]	60	28
249	3	1109	1768	gi 143156	membrane bound protein [Bacillus subtilis]	60	41
251	3	4053	1933	gi 1235662	RfbC [Myxococcus xanthus]	60	42
256	4	2614	3867	gi 532612	ecotropic retrovirus receptor [Mus musculus]	60	37
260	2	1539	802	gi 1208447	metalloprotease transporter [Serratia marcescens]	60	35
261	5	4528	3179	gnl PID e246728	histidine kinase [Streptococcus gordonii]	60	25
269	3	2723	1563	gi 1591618	M. jannaschii predicted coding region MJ0951 [Methanococcus jannaschii]	60	39
269	4	3541	2780	gi 1303794	YqeM [Bacillus subtilis]	60	36
269	11	7164	6595	gi 1303787	YqeG [Bacillus subtilis]	60	38
271	2	677	1651	gnl PID e269877	riboflavin kinase [Bacillus subtilis]	60	43
271	3	1639	2247	gi 537148	ORF_f181 [Escherichia coli]	60	41
271	18	13502	13762	pir S39341 S39341	grpE protein - Lactococcus lactis similar to S. typhimurium apba [Escherichia coli]	60	40
277	2	1662	979	gi 1773109		60	41
279	13	10627	9773	gi 290545	f270 [Escherichia coli]	60	41
290	2	790	1695	gi 152886	elongation factor Ts (tsf) [Spiroplasma]	60	38

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
291	4	3571	2612	gnl PID e257610	citri  sugar-binding transport protein [Anaerocellum thermophilum]	60	40
295	3	1309	2094	gi 1000453	TreR [Bacillus subtilis]	60	37
301	15	11063	11344	gi 535274	ORF1 [Streptococcus thermophilus]	60	36
310	3	2903	1266	gi 809765	aspartate aminotransferase (AA 1-402) [Sulfolobus solfataricus] pir S07088 S07088 aspartate transaminase (EC 2.6.1.1) - Sulfolobus olfataricus	60	44
316	2	319	119	bbs 115298	polyprotein(coat protein) [raspberry ringspot virus RRV, Peptide, 1107 aa] [Raspberry ringspot virus]	60	28
320	4	3085	2483	gi 143002	proton glutamate symport protein [Bacillus caldotenax] pir S26246 S26246 glutamate/aspartate transport protein - Bacillus aldotenax	60	26
323	1	1	681	gi 1477486	transposase [Burkholderia cepacia]	60	44
330	4	3361	4488	gi 1778517	glycerol dehydrogenase homolog [Escherichia coli]	60	48
356	3	2471	2205	gi 57633	neuronal myosin heavy chain [Rattus rattus]	60	40
362	5	2458	2925	gnl PID e255090	hypothetical protein [Bacillus subtilis]	60	36
364	4	4096	5349	gi 1657522	hypothetical protein [Escherichia coli]	60	41
383	1	654	4	gnl PID e288399	F56H6.k [Caenorhabditis elegans]	60	39
383	2	2208	853	gi 143536	sigma factor 54 [Bacillus subtilis]	60	37
386	2	130	510	gi 1046053	hypothetical protein (SP:P32049) [Mycoplasma genitalium]	60	42
399	26	25892	27757	gi 895747	putative cel operon regulator [Bacillus subtilis]	60	30
399	27	27721	28239	gi 146281	gut operon activator (gutM) [Escherichia coli]	60	35
401	4	2081	3523	gi 142833	ORF2 [Bacillus subtilis]	60	36

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
405	2	1353	763	gi 633113	ORF3 [Streptococcus sobrinus]	60	42
407	7	4380	4589	gi 1674126	(AE000043) Mycoplasma pneumoniae, MG280 homolog, from M. genitalium [Mycoplasma pneumoniae]	60	39
408	1	12	539	gi 455006	orf6 [Rhodococcus fascians]	60	42
421	7	4113	3925	gi 60020	ORF31 (AA1-868) [Human herpesvirus 3]	60	43
452	3	712	2223	gi 532554	ORF21 [Enterococcus faecalis]	60	38
462	3	2066	1551	gi 1015903	ORF YJR151c [Saccharomyces cerevisiae]	60	37
480	1	12	272	gi 468715	sss gene product [Pseudomonas aeruginosa]	60	34
487	1	1091	3	gi 388269	traC [Plasmid pAD1]	60	39
490	5	2108	1479	gi 699379	glvr-1 protein [Mycobacterium leprae]	60	29
507	1	221	751	gi 1303952	YqjA [Bacillus subtilis]	60	37
511	1	449	63	gi 391610	farnesyl diphosphate synthase [Bacillus stearothermophilus] pir JX0257 JX0257 geranyltransferase (EC 2.5.1.10) - Bacillus stearothermophilus	60	42
551	2	1521	604	gi 1256648	putative [Bacillus subtilis]	60	37
552	1	887	63	gi 537235	Kenn Rudd identifies as gpMB [Escherichia coli]	60	40
610	1	1	792	gi 1321625	exo-alpha-1, 4-glucosidase [Bacillus stearothermophilus]	60	45
642	1	402	214	gi 992964	thioredoxin [Arabidopsis thaliana]	60	36
646	1	642	265	gi 1041115	TRAC [Plasmid pPD1]	60	32
661	2	305	943	gi 1651536	3-oxoacyl-[acyl-carrier-protein] reductase [Escherichia coli]	60	37
678	1	536	3	gi 532554	ORF21 [Enterococcus faecalis]	60	39
716	1	799	305	gi 886040	ORFtxe1 [Clostridium difficile]	60	38
717	1	2	472	gi 1402529	ORF8 [Enterococcus faecalis]	60	31
727	1	516	82	gi 471283	ORF [Synechococcus PCC6301]	60	41
770	1	327	4	gi 467451	unknown [Bacillus subtilis]	60	33
843	1	234	4	gi 2819	transferase [GAL10] (AA 1 - 687)	60	37

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					[Kluyveromyces lactis] r[S01407]XUVKG UDPGlucose 4-epimerase (EC 5.1.3.2) - yeast uyveromyces marxianus var. lactis)		
21	1	341	3	gi 1778519	hypothetical protein [Escherichia coli]	59	47
23	2	290	1303	gi 1407800	ABC-type permease [Yersinia pestis]	59	36
23	13	6720	7388	gi 1652472	ethylene response sensor protein [Synecocystis sp.]	59	37
23	18	11892	12413	gi 825627	major carboxysome shell protein [Thiobacillus neapolitanus] pir S60136 S60136 major carboxysome shell protein - Thiobacillus eapolitanus	59	42
29	4	1989	2852	gi 1742383	ORF_ID:0276#3; similar to [PIR Accession Number S11432] [Escherichia coli]	59	48
32	8	4504	4064	gi 1046081	hypothetical protein (GB:D26185_10) [Mycoplasma genitalium]	59	33
37	9	6670	6284	gi 290561	o188 [Escherichia coli]	59	44
47	1	2	2743	gnl PID e248792	unknown [Mycobacterium tuberculosis]	59	46
48	5	4017	5492	gi 1185288	isochorismate synthase [Bacillus subtilis]	59	40
49	5	1797	2093	gi 496280	structural protein [Bacteriophage Tuc2009]	59	41
59	8	3324	5057	gi 1486244	unknown [Bacillus subtilis]	59	35
72	14	13937	13434	gi 532540	ORF7 [Enterococcus faecalis]	59	25
81	20	14659	14219	gi 39978	P16 [Bacillus subtilis]	59	38
98	2	1961	2617	gi 41519	P30 protein (AA 1-240) [Escherichia coli]	59	39
102	3	2542	3774	gi 1674376	(AE000062) Mycoplasma pneumoniae, MG148 homolog, from M. genitalium [Mycoplasma pneumoniae]	59	30
116	2	907	1458	gi 1146225	putative [Bacillus subtilis]	59	37
116	7	3532	4842	gi 1146238	poly(A) polymerase [Bacillus subtilis]	59	41
128	20	15626	14310	gi 1001719	ATP-dependent RNA helicase Dead [Synecocystis sp.]	59	34
134	4	3158	3850	gi 1477486	transposase [Burkholderia cepacia]	59	40

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
137	1	1	999	gi 1065948	similar to thymidine diphosphoglucose 4,6-dehydratase [Caenorhabditis elegans]	59	40
138	8	7489	6827	gnl PID e264435	Putative orf YCLX8c, len:192 [Saccharomyces cerevisiae]	59	36
140	1	3	656	gnl PID e254943	unknown [Mycobacterium tuberculosis]	59	32
165	13	10427	9849	gi 1732199	PTS permease for mannose subunit IIIMan C terminal domain [Vibrio furnissii]	59	37
167	1	2	1045	gi 1573128	hypothetical [Haemophilus influenzae]	59	38
173	2	430	2160	gi 1486244	unknown [Bacillus subtilis]	59	31
179	10	10432	11199	gi 288299	ORF1 gene product [Bacillus megaterium]	59	34
179	12	12117	13148	gi 1045964	hypothetical protein (GB:U14003_297) [Mycoplasma genitalium]	59	41
181	11	9684	8575	gi 1653152	3-dehydroquinate synthase [Synechocystis sp.]	59	41
223	24	20736	21974	gi 1573051	succinyl-diaminopimelate desuccinylase (dapE) [Haemophilus influenzae]	59	48
229	12	12818	11421	gi 1652035	fmu and fmv protein [Synechocystis sp.]	59	39
244	3	2836	1565	gi 1303959	YqjH [Bacillus subtilis]	59	45
265	9	4116	3868	gi 311100	translational activator [Saccharomyces cerevisiae]	59	28
272	1	1	546	gi 490320	Y gene product [unidentified]	59	41
279	16	14774	14370	gi 1389549	ORF3 [Bacillus subtilis]	59	46
283	8	3222	3401	gi 153047	lysostaphin (ttg start codon) [Staphylococcus simulans] pir A25881 A25881 lysostaphin precursor - Staphylococcus simulans sp P10547 LSTP_STASI LYOSTAPHIN PRECURSOR (EC 3.5.1.-).	59	43
288	5	2617	3144	gi 1142714	phosphoenolpyruvate:mannose phosphotransferase element IIB [Lactobacillus curvatus]	59	45
292	19	14837	16792	gi 495646	ATPase [Transposon Tn5422]	59	40

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
295	1	49	495	gi 533098	DnaD protein [Bacillus subtilis]	59	39
315	2	907	653	gi 1574802	hypothetical [Haemophilus influenzae]	59	38
318	6	4549	4058	gi 43941	EIII-B Sor PTS [Klebsiella pneumoniae]	59	35
345	3	2707	3507	gi 895749	putative cellobiose phosphotransferase enzyme II'' [Bacillus ubtilis]	59	38
351	5	2646	2371	gi 1666506	Rfbc [Leptospira interrogans]	59	30
355	21	15237	17222	gi 515738	ORF2; putative [Oenococcus oeni]	59	35
384	1	14	754	gi 1162959	homologous to HI0365 in Haemophilus influenzae; ORF1 [Pseudomonas aeruginosa]	59	34
385	1	3	533	gi 1146197	putative [Bacillus subtilis]	59	36
394	13	13137	12160	gnl PID e243582	ORF YGR263c [Saccharomyces cerevisiae]	59	36
399	1	224	580	gi 580904	homologous to E.coli rnpA [Bacillus subtilis]	59	38
412	1	3	2927	gi 1620648	surface protein Rib [Streptococcus agalactiae]	59	43
412	2	2918	3559	gi 1620648	surface protein Rib [Streptococcus agalactiae]	59	43
416	6	5283	3940	gi 1100076	PTS-dependent enzyme II [Clostridium longisporum]	59	38
437	2	1561	1136	gi 580866	ipa-12d gene product [Bacillus subtilis]	59	44
495	2	438	614	gi 1500472	M. jannaschii predicted coding region MJ1577 [Methanococcus jannaschii]	59	45
502	1	853	188	gi 1063248	No homologous protein [Bacillus subtilis]	59	25
573	8	5092	4493	gi 1573226	hypothetical [Haemophilus influenzae]	59	39
579	4	1716	2717	gnl PID e280724	unknown [Mycobacterium tuberculosis]	59	41
600	1	1	504	gi 49386	internal region of the penicillin-binding protein 2B gene treptococcus pneumoniae]	59	40
616	3	904	533	gi 289265	[Bacillus sp. (KSM 64) endo-1,4-beta-glucanase gene, complete cds.], ene products [Bacillus sp.]	59	44
657	1	432	4	gi 1651338	PnuC protein [Escherichia coli]	59	37

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
699	1	416	165	gnl PID e199096	PepR1 [Lactobacillus delbrueckii]	59	23
713	4	3709	2660	gi 515738	ORF2; putative [Oenococcus oeni]	59	37
715	1	698	84	gi 1176399	EpIF [Staphylococcus epidermidis]	59	42
737	2	660	199	gi 666000	hypothetical protein [Bacillus subtilis]	59	43
744	1	395	3	gi 1732057	MUC-CL-1 [Trypanosoma cruzi]	59	45
746	1	3	554	gi 141858	replication-associated protein [Plasmid pAD1]	59	36
869	1	2	250	gi 1432153	cellobiose-specific PTS permease [Klebsiella oxytoca]	59	40
4	8	6948	6067	gi 147516	ribokinase [Escherichia coli]	58	42
11	6	3312	4121	gi 1732200	PTS permease for mannose subunit IIPMan [Vibrio furnissii]	58	35
16	9	7684	6932	gnl PID e233879	hypothetical protein [Bacillus subtilis]	58	48
23	14	7440	8903	gi 142940	ftsA [Bacillus subtilis]	58	39
30	2	570	1283	gi 1644202	unknown [Bacillus subtilis]	58	37
48	7	7186	8037	gi 1573247	hypothetical [Haemophilus influenzae]	58	35
49	7	2395	2871	gnl PID e210884	c2 gene product [Bacteriophage B1]	58	34
54	1	1014	91	gi 46645	ORF (rlx) [Staphylococcus aureus]	58	46
55	3	1221	511	gi 726443	No definition line found [Caenorhabditis elegans]	58	41
58	1	1904	696	gi 1591564	molybdenum cofactor biosynthesis moeA protein [Methanococcus jannaschii]	58	39
58	8	7238	6996	gi 1279769	FdhC [Methanobacterium thermoformicum]	58	54
72	12	12117	10897	gi 763052	integrase [Bacteriophage T270]	58	37
77	2	1155	1910	gi 1245464	YfeA [Yersinia pestis]	58	34
78	1	2589	49	gi 40663	sialidase [Clostridium septicum]	58	40
88	9	5854	6528	gi 1619623	hemin binding protein [Yersinia enterocolitica]	58	37
93	6	2639	2863	gi 405133	putative [Bacillus subtilis]	58	33
98	13	13523	12432	gi 147329	transport protein [Escherichia coli]	58	41
100	12	8550	8224	gi 1736642	Invasin. [Escherichia coli]	58	47

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
102	7	5688	5969	gi 808869	human gcp372 [Homo sapiens]	58	30
105	5	3716	4501	gi 143729	transcription activator [Bacillus subtilis]	58	40
107	1	511	2	gi 1303827	Yqfi [Bacillus subtilis]	58	34
108	2	1040	1732	gi 1592142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	58	37
114	6	7608	8444	gi 152719	flavocytochrome c [Shewanella putrefaciens]	58	40
117	14	11813	11115	gi 1575577	DNA-binding response regulator [Thermotoga maritima]	58	42
122	1	1	936	gi 393269	adhesion protein [Streptococcus pneumoniae]	58	38
123	23	20379	21617	gi 1653948	hypothetical protein [Synecocystis sp.]	58	38
133	8	7362	8480	gi 143498	degS protein [Bacillus subtilis]	58	38
133	9	8437	9087	gi 143089	iep protein [Bacillus subtilis]	58	31
138	3	3551	2898	gi 216114	DNA polymerase [Bacteriophage SP01]	58	41
138	5	5819	5049	gnl PID e289148	highly similar to phosphotransferase system regulator [Bacillus subtilis]	58	38
138	17	11419	10379	gi 1674137	(AE000044) Mycoplasma pneumoniae, lipocate protein ligase; similar to Swiss-Prot Accession Number P32099, from E. coli [Mycoplasma pneumoniae]	58	37
139	8	5002	4808	gi 153607	dpsD gene product [Streptococcus pneumoniae]	58	43
146	9	7817	6627	gi 606076	ORF_0384 [Escherichia coli]	58	43
150	10	7529	7894	gi 141852	sialidase [Actinomyces viscosus]	58	28
152	10	5717	6637	gi 296356	putative membrane transport protein [Clostridium perfringens]	58	36
162	10	11009	11185	gi 42655	pir A56641 A56641 probable membrane transport protein - Clostridium perfringens	58	37
164	3	1793	1608	gi 881499	pi protein [Escherichia coli]	58	41
					parathion hydrolase (phosphotriesterase)-	58	



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
165	6	5640	4975	gi 1146190	related protein [Mus usculus]		
165	10	9038	8199	gi 606080	2-keto-3-deoxy-6-phosphogluconate aldolase [Bacillus subtilis]	58	39
168	1	1	657	gi 413930	ORF_0290; Geneplot suggests frameshift linking to o267, not found Escherichia coli	58	35
170	1	923	234	gi 1573505	ipa-6d gene product [Bacillus subtilis]	58	41
176	1	1	1101	gi 1652379	hypothetical [Haemophilus influenzae]	58	30
180	12	10237	10410	gi 408123	cation-transporting P-ATPase [Synechocystis sp.]	58	30
193	3	2077	1388	gi 1256633	V-ATPase 14kD subunit peptide [Drosophila melanogaster] pir S38436 S38436 H+-transporting ATPase (EC 3.6.1.35) 14K chain - fruit fly (Drosophila melanogaster)	58	33
193	4	2602	2075	gi 147920	putative [Bacillus subtilis]	58	39
194	9	6492	5500	sp P09997 YIDA_ECO LI	3-methyladenine-DNA glycosylase I (tag) [Escherichia coli]	58	33
201	5	5152	4466	gi 755152	HYPOTHETICAL 29.7 KD PROTEIN IN IBPA-GYRB INTERGENIC REGION.	58	38
210	9	6546	7265	gi 466520	highly hydrophobic integral membrane protein [Bacillus subtilis]	58	28
220	1	3	569	gi 467441	sp P42953 TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG.	58	36
222	10	6520	7143	gi 1674024	pocr [Salmonella typhimurium]	58	38
					expressed at the end of exponential growth under conditions in which he enzymes of the TCA cycle are repressed [Bacillus subtilis] sp P14194 CTC_BACSU GENERAL STRESS PROTEIN CTC. (SUB 2-204) gi 40219 partial ctc gene product (AA 1-186) [Bacillus subtilis]		
					(AE000033) Mycoplasma pneumoniae,	58	41

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					hypothetical protein (yjfS) homolog; similar to Swiss-Prot Accession Number P39301, from <i>E. coli</i> [Mycoplasma pneumoniae]		
233	7	4984	3944	gi 147806	selenium metabolism protein [Escherichia coli]	58	45
238	14	12128	12910	gi 1736468	Pectin degradation repressor protein Kdgr. [Escherichia coli]	58	37
244	11	8102	7809	gi 467418	unknown [Bacillus subtilis]	58	37
246	1	1	276	gi 65291	receptor tyrosine kinase preprotein (Xiphophorus sp.) ir S06142 S06142 kinase-related transforming protein (Tu) (EC 7.1.-) precursor - southern platyfish	58	32
255	4	2927	2559	gi 1652384	ABC transporter [Synechocystis sp.]	58	41
258	9	8025	8966	gi 147402	mannose permease subunit III-Man [Escherichia coli]	58	35
259	2	1801	893	gi 1591564	molybdenum cofactor biosynthesis moeA protein [Methanococcus jannaschii]	58	39
260	3	1754	2254	gi 580841	F1 [Bacillus subtilis]	58	38
271	4	2382	2738	gi 40067	X gene product [Bacillus sphaericus]	58	37
279	8	6237	6536	gi 1783243	homologous to jojC gene product (B. subtilis; prf:211327a); hypothetical [Bacillus subtilis]	58	34
301	1	753	175	gi 499196	ORF1 [Streptomyces lincolnensis]	58	37
304	1	100	849	gi 1653322	hypothetical protein [Synechocystis sp.]	58	41
313	2	748	1650	gi 1658371	cyclic beta-1,2-glucan modification protein [Rhizobium meliloti]	58	36
321	11	6033	6533	gi 1573292	hypothetical [Haemophilus influenzae]	58	34
322	6	3819	5069	gi 23897	5'-nucleotidase [Homo sapiens]	58	34
324	5	3259	4452	gi 1469784	putative cell division protein ftsw [Enterococcus hirae]	58	37
328	1	1	270	gi 882579	CG Site No. 29739 [Escherichia coli]	58	43

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
330	8	6228	6758	gi 43941	EIII-B Sor PTS [Klebsiella pneumoniae]	58	37
334	4	3634	3963	gi 1001306	hypothetical protein [Synecocystis sp.]	58	34
345	17	18899	20044	gi 853809	ORF3 [Clostridium perfringens]	58	30
363	7	8475	9944	gi 348056	trans-acting positive regulator [Bacillus anthracis]	58	33
375	7	6472	5279	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	58	42
394	12	10689	12095	gi 537034	ORF_o488 [Escherichia coli]	58	32
399	3	1383	2198	gi 580905	B. subtilis genes rpmH, rnpA, 50kd, gida and gidB [Bacillus subtilis] gi 580919 Jag [Bacillus subtilis]	58	36
399	16	11544	12098	gi 1572965	hypothetical [Haemophilus influenzae]	58	39
399	19	14776	15654	gi 1778530	CitG homolog [Escherichia coli]	58	40
407	2	738	553	gi 170553	pyruvate kinase [Trichoderma reesei]	58	38
416	5	4045	3389	gi 475112	enzyme IIabc [Pediococcus pentosaceus]	58	41
449	4	1421	879	gi 928834	integrase [Lactococcus lactis phage BK5-T]	58	32
497	1	3	458	gi 160628	reticulocyte binding protein 2 [Plasmodium vivax]	58	30
594	1	285	4	gi 1353874	unknown [Rhodobacter capsulatus]	58	39
637	6	3451	2765	pir D61615 D61615	sericin MG-1 - greater wax moth (fragment)	58	52
653	1	595	245	gi 1408585	LtrD [Lactococcus lactis lactis]	58	41
656	4	3713	5209	sp P13692 P54_ENTF_C	P54 PROTEIN PRECURSOR.	58	37
656	6	5988	6467	gi 1017818	phosphotyrosine protein phosphatase [Streptomyces coelicolor]	58	48
667	1	88	1467	bbs 177441	OsNramp1=Nramp1 homolog/Bcg product homolog [Oryza sativa, indica, cv. IR 36, etiolated shoots, Peptide, 517 aa] [Oryza sativa]	58	40
686	1	892	233	pir A24255 A24255	chorion class A protein L11 precursor -	58	38

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
					silkworm		
706	1	1002	607	gi 1001762	hypothetical protein [Synecocystis sp.]	58	32
801	1	254	12	gnl PID e243641	unknown [Mycobacterium tuberculosis]	58	29
848	1	212	3	gnl PID e254644	membrane protein [Streptococcus pneumoniae]	58	37
975	1	3	422	gi 290545	f270 [Escherichia coli]	58	35
11	4	2345	2833	gi 1439527	EITA-man [Lactobacillus curvatus]	57	46
16	2	1426	365	gi 780550	acetyl transferase [Rhizobium loti]	57	35
18	3	1593	925	gnl PID e137594	xerC recombinase [Lactobacillus leichmannii]	57	36
19	15	8058	8267	gi 1590922	cell division inhibitor [Methanococcus jannaschii]	57	42
19	23	11938	12318	gi 1294760	structural protein; orfL3; putative [Bacteriophage phi-41]	57	46
25	9	7743	6958	gnl PID e255000	hypothetical protein [Bacillus subtilis]	57	40
47	3	3857	4462	gi 1353540	ORF23 [Bacteriophage rlt]	57	35
65	10	7180	8919	gi 496254	fibronectin/fibrinogen-binding protein [Streptococcus pyogenes]	57	40
68	7	3923	3705	gi 336656	ribosomal protein secY [Cyanophora paradoxa]	57	28
70	4	2317	3645	pir S11158 YESAEE	erythromycin resistance protein - Staphylococcus epidermidis plasmid pUL5050	57	40
76	1	55	1095	gi 1353562	Structural protein [Bacteriophage rlt]	57	41
91	11	9070	8849	gi 550321	beta-fructofuranosidase [Chenopodium rubrum]	57	30
94	4	1740	1495	gi 47406	penicillin-binding protein 1a [Streptococcus pneumoniae] ir S28031 S28031 penicillin-binding protein 1a - Streptococcus pneumoniae (strain 456) (fragment)	57	30
98	6	7766	6849	gi 409286	bmrU [Bacillus subtilis]	57	31

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
100	22	17294	15912	gnl PID e289150	member of the SNF2 helicase family [Bacillus subtilis]	57	30
102	1	66	2465	gi 405564	traE [Plasmid pSK41]	57	28
110	14	11757	12497	gi 854601	unknown [Schizosaccharomyces pombe]	57	38
114	9	10291	11139	gi 853777	product similar to E.coli PRFA2 protein [Bacillus subtilis] pir S55438 S55438 ywKE protein - Bacillus subtilis sp P45873 HEMK_BACSU POSSIBLE PROTOPORPHYRINOGEN OXIDASE (EC .3.3.-).	57	38
115	3	955	1461	gi 396347	alternate name yjaB [Escherichia coli]	57	33
123	3	1925	2932	gi 1001731	low affinity sulfate transporter [Synechocystis sp.]	57	39
124	7	6026	5118	gi 1674310	(AE000058) Mycoplasma pneumoniae, MG085 homolog, from M. genitalium [Mycoplasma pneumoniae]	57	30
128	9	7530	6235	gi 413940	ipa-16d gene product [Bacillus subtilis]	57	36
128	31	25487	25206	gi 1651915	hypothetical protein [Synechocystis sp.]	57	42
128	33	26878	26150	gi 1001387	hypothetical protein [Synechocystis sp.]	57	30
128	37	30730	29600	gi 406877	DivIB protein [Bacillus licheniformis]	57	35
130	9	7408	8556	gi 343539	NADH dehydrogenase subunit 4 [Trypanosoma brucei]	57	27
144	1	1013	219	gi 1652518	hypothetical protein [Synechocystis sp.]	57	45
144	6	4145	5254	gi 149581	maturation protein [Lactobacillus paracasei]	57	38
146	1	617	192	gi 147402	mannose permease subunit III-Man [Escherichia coli]	57	33
153	1	83	991	gi 147336	transmembrane protein [Escherichia coli]	57	33
160	8	4718	4134	gi 305333	zeta-crystallin [Cavia porcellus]	57	39
167	8	14891	14688	gi 206354	protein kinase C, zeta subspecies [Rattus norvegicus] pir A30314 A30314 protein kinase C (EC 2.7.1.-) zeta - rat sp P09217 KPCZ_RAT PROTEIN KINASE C, ZETA	57	39

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
174	1	760	2	gnl PID e191403	TYPE (EC 2.7.1.1.-) NPKC-ZETA). ORFA gene product [Chloroflexus aurantiacus]	57	42
176	4	3347	3568	gi 1236529	cyclonaltodextrinase [Bacillus sp.]	57	46
194	8	4786	5457	gi 405516	This ORF is homologous to nitroreductase from Enterobacter cloacae, ccession Number A38686, and Salmonella, Accession Number P15888 Mycoplasma-like organism]	57	26
199	3	3207	3764	gi 216350	ORF [Bacillus subtilis]	57	38
202	5	3356	3664	gi 1183841	Holliday junction binding protein [Pseudomonas aeruginosa]	57	34
202	12	10911	10192	gi 971338	anaerobic regulatory protein [Bacillus subtilis]	57	27
205	3	1022	468	gi 1783240	hypothetical [Bacillus subtilis]	57	38
223	2	779	1501	gi 1208965	hypothetical 23.3 kd protein [Escherichia coli]	57	32
223	3	1499	2332	gi 303560	ORF271 [Escherichia coli]	57	35
223	11	8404	12198	gi 158079	period protein [Drosophila serrata]	57	40
237	9	3685	3906	gi 514919	phosphofructokinase [Drosophila melanogaster]	57	31
242	7	5760	5020	gi 1574596	H. influenzae predicted coding region HI1738 [Haemophilus influenzae]	57	33
250	2	1243	1485	gnl PID e275819	K0862.8 [Caenorhabditis elegans]	57	47
276	28	16565	16332	gi 886375	variant-specific surface protein [Plasmodium falciparum]	57	47
288	6	3157	3363	gi 147403	mannose permease subunit II-P-Man [Escherichia coli]	57	39
289	1	141	818	gi 1742822	Phosphoglycolate phosphatase (EC 3.1.3.18). [Escherichia coli]	57	40
292	20	15930	15721	gi 854201	putative polymerase [infectious bursal disease virus]	57	47
294	4	1454	2014	gi 454303	LDV2 gene product [Allium porrum]	57	41

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
295	4	2052	2342	pir S48588 S48588	hypothetical protein - Mycoplasma capricolum (SGC3) (fragment)	57	39
301	14	10921	10148	gnl PID e262045	putative orf [Bacillus subtilis]	57	38
306	1	2	793	gi 216715	HpaI methyltransferase [Haemophilus parainfluenzae] pir S28681 S28681 site-specific DNA-methyltransferase adenine-specific (EC 2.1.1.72) HpaI - Haemophilus parainfluenzae sp P29538 MTH1_HAEPA MODIFICATION METHYLASE HPAI (EC 2.1.1.72) ADENINE-SPECIFIC MET	57	36
306	8	5418	5663	gi 1591542	M. jannaschii predicted coding region MJ0857 [Methanococcus jannaschii]	57	42
308	2	1732	1487	gi 1518045	FlbF protein [Borrelia burgdorferi]	57	28
321	2	1030	1458	gi 606080	ORF_o290; Geneplot suggests frameshift linking to o267, not found Escherichia coli	57	30
351	4	2342	1587	gi 1591853	M. jannaschii predicted coding region MJ1222 [Methanococcus jannaschii]	57	37
355	30	20619	20861	gi 1136394	There are three putative hydrophobic domains in the central region. [Homo sapiens]	57	42
364	10	9415	8852	gi 38722	precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir A29277 A29277 aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter lcoaceticus	57	32
365	3	4715	1812	gi 914990	Similar to DEAD box family helicases [Saccharomyces cerevisiae] pir S59797 S59797 hypothetical protein D9798.1 - yeast Saccharomyces cerevisiae	57	35
378	1	615	10	gi 1652989	hypothetical protein [Synecocystis sp.]	57	35
379	1	1457	114	gi 1256618	transport protein [Bacillus subtilis]	57	36
390	1	1426	2	gi 387880	collagen adhesin [Staphylococcus aureus]	57	37

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
422	1	2	409	gi 1591837	M. jannaschii predicted coding region MJ1207 [Methanococcus jannaschii]	57	37
447	1	397	131	gi 214566	keratin protein XK81 [Xenopus laevis]	57	33
454	2	1095	889	gi 1783256	sigma factor [Bacillus subtilis]	57	28
504	2	641	1426	gi 42081	nagD gene product (AA 1-250) [Escherichia coli]	57	32
524	2	963	577	gi 143724	putative [Bacillus subtilis]	57	43
535	4	4862	4305	gi 146549	kdpC [Escherichia coli]	57	40
547	2	426	719	gi 533098	DnaD protein [Bacillus subtilis]	57	33
548	1	316	717	gi 397973	Mg2+ transport ATPase [Salmonella typhimurium]	57	33
639	2	359	105	gnl PID e247390	P-type ATPase [Dictyostelium discoideum]	57	31
641	1	941	180	gnl PID e261990	putative orf [Bacillus subtilis]	57	36
686	3	1298	3259	gi 496506	orf gamma [Streptococcus pyogenes]	57	37
686	6	2200	2847	gi 404800	putative [Saccharopolyspora erythraea]	57	47
782	2	591	860	gi 1591270	alanyl-tRNA synthetase [Methanococcus jannaschii]	57	32
844	1	3	182	gi 849217	Weak similarity to Streptococcus Protein V, a type-II IgG receptor PIR accession number S17354) and Giardia lamblia median body rotein (PIR accession number S33821) [Saccharomyces cerevisiae] pir S61181 S61181 hypothetical protein D9740.10 - yeast Sacchar	57	34
859	1	174	4	gi 1762584	polygalacturonase isoenzyme 1 beta subunit homolog [Arabidopsis thaliana]	57	28
967	1	381	4	gi 309662	pheromone binding protein [Plasmid pCF10]	57	40
11	5	2817	3314	gi 43941	EIII-B Sor PTS [Klebsiella pneumoniae]	56	30
15	1	80	892	gi 1574803	spermidine/putrescine-binding periplasmic protein precursor (potD) [Haemophilus influenzae]	56	32
37	8	6327	6088	gi 290561	ol88 [Escherichia coli]	56	41



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
44	2	1169	1360	gi 16096	peroxidase [Armoracia rusticana]	56	37
56	3	1881	1363	gi 49272	Asparaginase [Bacillus licheniformis]	56	33
65	1	102	887	gi 1377832	unknown [Bacillus subtilis]	56	41
75	9	5817	4306	gi 1235712	polyprotein [Infectious pancreatic necrosis virus]	56	30
83	7	3260	4051	gi 1652645	phosphoglycolate phosphatase [Synechocystis sp.]	56	30
95	3	1793	2389	pir C53610 C53610	ntpE protein - Enterococcus hirae	56	28
100	3	5076	1915	gi 1353559	ORF42 [Bacteriophage rlt]	56	35
100	16	10581	10369	gi 868224	No definition line found [Caenorhabditis elegans]	56	35
100	48	31841	32770	gi 460025	ORF2, putative [Streptococcus pneumoniae]	56	38
108	5	4007	3336	gi 288301	ORF2 gene product [Bacillus megaterium]	56	34
109	2	1032	325	gi 413976	ipa-52r gene product [Bacillus subtilis]	56	36
119	7	3958	5304	gi 498842	VirS [Clostridium perfringens]	56	35
123	32	29479	30345	gi 39981	P30 [Bacillus subtilis]	56	38
126	1	521	3	gi 147403	mannose permease subunit II-P-Man [Escherichia coli]	56	29
130	6	4296	6104	gi 308854	oligopeptide binding protein [Lactococcus lactis]	56	33
131	7	5267	6613	gi 466589	CG Site No. 39 [Escherichia coli]	56	32
133	5	4358	5758	gi 1573431	aminodeoxychorismate lyase (pabC) [Haemophilus influenzae]	56	40
138	20	13680	12670	gi 1590951	UDP-glucose 4-epimerase [Methanococcus jannaschii]	56	40
138	29	19764	18823	gi 44864	H.8 outer membrane protein (AA -17 to 71) [Neisseria gonorrhoeae] ir S02720 S02720 outer membrane protein H.8 precursor - Neisseria norrhoeae	56	33
145	7	5611	7179	gi 1652892	ABC transporter [Synechocystis sp.]	56	33
146	10	8545	7811	gi 41519	P30 protein (AA 1-240) [Escherichia coli]	56	28

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
150	4	2979	4637	gi 309662	pheromone binding protein [Plasmid pCF10]	56	32
159	5	5362	5066	gi 576733	apocytochrome b [Trypanoplasma borreli]	56	43
164	13	8864	15031	gi 1654116	protein F2 [Streptococcus pyogenes]	56	43
179	7	7790	9118	gi 413926	ipa-2r gene product [Bacillus subtilis]	56	33
187	4	2239	1667	gi 1573061	hypothetical [Haemophilus influenzae]	56	18
200	19	11473	10724	gi 498817	ORF8; homologous to small subunit of phage terminases [Bacillus ubtilis]	56	35
206	6	3766	2759	gi 474837	ORF1 [Thermoanaerobacterium thermosulfurigenes] sp P38541 YAMB_THETU HYPOTHETICAL 35.6 KD PROTEIN IN AMYB 5'REGION ORF1).	56	34
207	2	2091	1672	gi 1204258	soluble protein [Escherichia coli]	56	40
217	9	6661	6158	gi 1017427	elastic titin [Homo sapiens]	56	28
225	7	6007	5099	gi 1742675	Phosphotransferase system enzyme II (EC 2.7.1.69) MalX [Escherichia coli]	56	46
230	3	595	3153	gi 437706	alternative truncated translation product from E.coli [Streptococcus pneumoniae]	56	34
236	2	1486	515	gi 415664	catabolite control protein [Bacillus megaterium] sp P46828 CCPA_BACME GLUCOSE-RESISTANCE AMYLASE REGULATOR CATABOLITE CONTROL PROTEIN).	56	35
236	7	9255	8599	gi 343544	ATPase 6 [Trypanosoma brucei]	56	48
238	15	13059	13718	gi 1146190	2-keto-3-deoxy-6-phosphogluconate aldolase [Bacillus subtilis]	56	37
238	20	17734	18756	gi 1574060	hypothetical [Haemophilus influenzae]	56	32
238	23	21613	20726	gi 151361	member of the Arac/XylS family of transcriptional regulators Pseudomonas aeruginosa]	56	36
242	6	4103	4477	gi 886858	nicotinic acetylcholine receptor [Caenorhabditis elegans] pir S57648 S57648 nicotinic acetylcholine receptor - Caenorhabditis legans	56	35

Table 3: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
260	5	3170	3781	gnl PID e58151	F3 [Bacillus subtilis]	56	43
279	6	5140	2831	gi 581100	gamma-glutamylcysteine synthetase (aa 1-518) [Escherichia coli] pir A24136 SYECEC glutamate--cysteine ligase (EC 6.3.2.2) - scherichia coli	56	42
279	9	6434	7228	gi 1783243	homologous to joJc gene product (B. subtilis; prf:2111327a); hypothetical [Bacillus subtilis]	56	29
292	14	10719	11504	gi 45738	ORFC [Enterococcus faecalis]	56	37
313	3	3039	1831	gi 474915	orf 337; translated orf similarity to SW: BCR_ECOLI bicyclomycin esistance protein of Escherichia coli [Coxiella burnetii] pir S44207 S44207 hypothetical protein 337 - Coxiella burnetii (SUB -338)	56	31
313	5	4233	3589	gi 405883	yeiL [Escherichia coli]	56	30
322	5	1994	3715	gi 1377831	unknown [Bacillus subtilis]	56	34
353	2	2353	1310	gnl PID e254644	membrane protein [Streptococcus pneumoniae]	56	26
394	14	13289	14143	gi 142836	repressor protein [Bacillus subtilis]	56	30
399	32	30208	30891	gi 396293	similar to Bacillus subtilis hypoth. 20 kDa protein, in tsr 3' egion [Escherichia coli]	56	38
402	2	1267	914	gi 170710	alpha-type gliadin precursor protein [Triticum aestivum]	56	45
408	4	2825	2220	gnl PID e257696	collagen binding protein [Lactobacillus reuteri]	56	36
432	5	3105	3302	gi 11678	atpE gene product [Marchantia polymorpha]	56	33
443	2	844	1089	gi 1256138	ybbI [Bacillus subtilis]	56	36
499	2	875	1666	gi 1499876	magnesium and cobalt transport protein [Methanococcus jannaschii]	56	30
510	6	3864	4733	gi 147404	mannose permease subunit II-M-Man [Escherichia coli]	56	34

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
543	6	3706	3113	gi 563812	XCAP-C [Xenopus laevis]	56	32
609	2	390	653	gi 48745	principal sigma subunit (AA 1-442) [Streptomyces coelicolor] ir S11712 S11712 translation initiation factor sigma hrdB - reptomycetes coelicolor	56	37
626	2	1124	2104	gi 950197	unknown [Corynebacterium glutamicum]	56	40
787	1	2	634	gnl PID e283826	orf c04012 [Sulfolobus solfataricus]	56	26
820	1	1220	3	gi 44001	galactose-1-P-uridyl transferase [Lactobacillus helveticus] ir B47032 B47032 galactose-1-phosphate uridyl transferase - ctobacillus helveticus	56	35
875	1	1	144	gi 455178	16K protein [Escherichia coli]	56	46
906	2	307	846	gi 144858	ORF A [Clostridium perfringens]	56	34
941	1	3	335	gi 160299	glutamic acid-rich protein [Plasmodium falciparum] pir A54514 A54514 glutamic acid-rich protein precursor - Plasmodium alciparum	56	23
5	5	2451	2951	gi 1303811	YqeU [Bacillus subtilis]	55	39
8	10	8312	7947	gi 1196907	daunorubicin resistance protein [Streptomyces peucetius]	55	29
17	24	23626	24465	gnl PID e285322	RecX protein [Mycobacterium smegmatis]	55	28
17	31	31027	30344	gi 143830	xpaC [Bacillus subtilis]	55	22
17	34	31991	32302	gnl PID e229183	C11G6.3 [Caenorhabditis elegans]	55	34
30	1	2	478	pir S10655 S10655	hypothetical protein X - Pyrococcus woesei (fragment)	55	34
49	14	9998	10411	gi 455154	ORF D [Clostridium perfringens]	55	36
54	3	955	1332	gnl PID e238660	hypothetical protein [Bacillus subtilis]	55	32
54	10	3527	3231	pir JQ0405 JQ0405	hypothetical 119.5K protein (uvrA region) - Micrococcus luteus	55	45
67	4	2313	3044	gi 555750	unknown [Neisseria gonorrhoeae]	55	42
69	4	2250	2020	gnl PID e259955	K04G11.5 [Caenorhabditis elegans]	55	33

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
77	5	3954	2938	gi 1001634	hypothetical protein [Synechocystis sp.]	55	34
80	4	4806	2482	gi 465952	B1620_F1_30 [Mycobacterium leprae]	55	35
81	6	4212	3730	gi 606073	ORF_o169 [Escherichia coli]	55	34
83	1	66	737	gi 216064	morphogenesis protein B [Bacteriophage PZA]	55	36
89	10	9486	7714	gi 148221	DNA-dependent ATPase, DNA helicase [Escherichia coli] pir JS0137 BVECRQ recQ protein - Escherichia coli	55	35
91	5	2507	3289	gi 153015	FemA protein [Staphylococcus aureus]	55	35
100	14	9974	9393	gi 558603	synaptonemal complex protein 1 [Mus musculus]	55	30
116	1	1	909	gi 473901	ORF1 [Lactococcus lactis]	55	33
122	3	1801	2655	gi 1016216	putative protein of 299 amino acids [Cyanophora paradoxa]	55	28
123	30	28191	28721	gi 1142714	phosphoenolpyruvate:mannose phosphotransferase element IIB [Lactobacillus curvatus]	55	29
128	22	16664	16029	gi 606025	ORF_o221 [Escherichia coli]	55	42
150	7	5949	6521	gi 39573	P20 (AA 1-178) [Bacillus licheniformis]	55	32
155	7	5767	6660	gi 1763974	DPPA [Bacillus methanolicus]	55	31
157	1	867	70	gi 1067010	M153.1 [Caenorhabditis elegans]	55	34
160	9	6090	4804	gi 1592141	M. jannaschii predicted coding region MJ1507 [Methanococcus jannaschii]	55	31
176	3	2060	3349	gi 153858	wall-associated protein [Streptococcus mutans]	55	37
201	2	3277	413	gi 1235662	RfbC [Myxococcus xanthus]	55	36
202	9	6199	8001	gi 606018	ORF_o783 [Escherichia coli]	55	42
222	7	4803	4021	gnl PID e289148	highly similar to phosphotransferase system regulator [Bacillus subtilis]	55	40
238	12	11465	9942	gnl PID e266573	unknown [Mycobacterium tuberculosis]	55	27
238	13	11527	12027	gi 1129093	unknown protein [Bacillus sp.]	55	36

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
240	4	1988	1215	gnl PID e252616	DcuC protein [Escherichia coli]	55	34
246	2	433	792	gnl PID e233868	hypothetical protein [Bacillus subtilis]	55	25
253	5	1827	1549	gi 142540	aspartokinase II [Bacillus sp.]	55	48
259	1	895	74	gi 1006621	molybdate-binding periplasmic protein [Synechocystis sp.]	55	37
267	1	1183	2	gi 882672	ORF_0313 [Escherichia coli]	55	27
292	16	12843	13325	gi 561746	cyclin-dependent protein kinase [Mus musculus]	55	26
294	9	3390	3752	gi 984582	DinJ [Escherichia coli]	55	26
300	5	3914	3582	gi 1591957	M. jannaschii predicted coding region MJ1318 [Methanococcus jannaschii]	55	38
305	3	2769	3527	gi 606309	ORF_0265; gtg start [Escherichia coli]	55	36
320	6	4479	3475	gi 1591732	cobalt transport ATP-binding protein O [Methanococcus jannaschii]	55	32
355	24	18149	18322	gi 344751	MDV TK gene product [unidentified]	55	40
364	2	2083	386	gi 1573045	hypothetical [Haemophilus influenzae]	55	40
364	9	8796	8575	gnl PID e252108	ORF YOR255w [Saccharomyces cerevisiae]	55	27
379	8	8248	6872	gi 1330236	dihydropyrimidinase [Homo sapiens]	55	37
386	6	3847	4332	gi 976025	HrsA [Escherichia coli]	55	27
441	2	939	1730	gi 144859	ORF B [Clostridium perfringens]	55	28
482	6	3515	3156	gi 606162	ORF_f229 [Escherichia coli]	55	39
497	9	4885	5937	gi 1041637	replication initiator protein [Staphylococcus xylosum]	55	33
546	1	1	1104	gi 467446	similar to SpoVB [Bacillus subtilis]	55	36
634	4	2132	1524	gi 431950	similar to a B.subtilis gene (GB: BACHEMEHY_5) [Clostridium astreum]	55	27
660	2	249	401	gnl PID e254995	hypothetical protein [Bacillus subtilis]	55	35
671	1	288	58	gi 38722	precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir A29277 A29277 aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter lcoaceticus	55	33

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
686	2	245	1141	gi 1633572	Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus]	55	36
713	3	2742	1438	gnl PID e8901	RESA NF7 Ag13 [Plasmodium falciparum]	55	25
815	1	2	226	gi 1113815	histidine kinase [Borrelia burgdorferi]	55	36
857	1	2	520	gi 143024	glucose-resistance amylase regulator [Bacillus subtilis] pir S15318 S15318 ccpA protein - Bacillus subtilis sp P25144 CCPA_BACSU GLUCOSE-RESISTANCE AMYLASE REGULATOR CATABOLITE CONTROL PROTEIN)	55	31
931	1	3	557	gi 1098508	putative spore germination apparatus protein [Bacillus megaterium]	55	32
17	7	6379	7218	gnl PID e250887	potential coding region [Clostridium difficile]	54	35
21	9	7265	6348	gi 13441	NADH dehydrogenase subunit 4L [Phoca vitulina]	54	29
28	2	2727	3425	gi 1001792	hypothetical protein [Synechocystis sp.]	54	29
32	6	4044	3523	gi 1673660	(AE000002) Mycoplasma pneumoniae, hypothetical 28K protein; similar to GenBank Accession Number JS0068, from M. pneumoniae [Mycoplasma pneumoniae]	54	36
33	3	2274	3767	gnl PID e245024	unknown [Mycobacterium tuberculosis]	54	36
40	1	1	915	gi 773349	BirA protein [Bacillus subtilis]	54	32
49	6	2120	2485	gnl PID e139446	a2 gene product [Bacteriophage B1]	54	38
54	17	8969	8661	gi 334068	ORF2 [Suid herpesvirus 1]	54	51
65	2	1311	2120	gi 537207	ORF_f277 [Escherichia coli]	54	27
72	20	21986	22435	gi 928848	ORF70'; putative [Lactococcus lactis phage BK5-T]	54	34
105	4	3039	3827	gnl PID e205174	orf2 gene product [Lactobacillus helveticus]	54	30
127	1	884	150	gi 726443	No definition line found [Caenorhabditis]	54	31

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
148	1	1204	62	gi 467456	elegans]		
156	4	4360	3167	gi 1032483	unknown [Bacillus subtilis]	54	37
160	4	1523	2077	gnl PID e255111	unidentified ORF downstream of hydrogenase cluster; ORF5 [Anabaena variabilis]	54	30
160	7	4260	3745	gi 1184121	hypothetical protein [Bacillus subtilis]	54	27
165	5	4996	3971	gi 1772652	auxin-induced protein [Vigna radiata]	54	30
176	2	1044	1937	gi 162201	2-keto-3-deoxygluconate kinase [Haloferax alicantei]	54	36
180	29	30833	29853	gnl PID e254644	P-type ATPase [Trypanosoma brucei]	54	38
200	16	7933	6656	gi 1574238	membrane protein [Streptococcus pneumoniae]	54	29
206	1	232	2	gi 1220501	tran protein (tran) [Haemophilus influenzae]	54	31
220	4	5235	4342	gi 606080	Rickettsia tsutsugamushi (strain Kp47) gene, complete cds [Rickettsia tsutsugamushi]	54	31
220	5	5821	5135	gi 43942	ORF_o290; Geneplot suggests frameshift linking to o267, not found Escherichia coli]	54	36
223	20	17253	17747	gi 47932	first subunit of EII-Sor [Klebsiella pneumoniae]	54	38
228	7	4866	4033	gi 1736828	tonB protein [Salmonella typhimurium]	54	34
229	4	5050	3371	gi 1046078	Thi4 protein [Escherichia coli]	54	42
236	3	4777	1496	gi 152271	M. genitalium predicted coding region MG369 [Mycoplasma genitalium]	54	28
236	5	7822	6944	gnl PID e285031	319-kDa protein [Rhizobium meliloti]	54	20
238	30	27964	27746	gnl PID e217586	Hyp1 protein [Hydra vulgaris]	54	42
242	5	3508	4050	gi 149502	PlnM [Lactobacillus plantarum]	54	35
257	1	296	120	gi 1498064	beta-lactamase [Lactococcus lactis]	54	50
257	6	6745	5633	gi 343949	AtE1 [Arabidopsis thaliana]	54	42
257	6	6745	5633	gi 343949	var1(40.0) [Saccharomyces cerevisiae]	54	42



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
258	8	7839	7114	gi 41519	P30 protein (AA 1-240) [Escherichia coli]	54	31
276	20	13101	12880	gi 155322	icsB gene product [Plasmid pWR100]	54	37
280	1	618	106	gi 467356	unknown [Bacillus subtilis]	54	21
288	4	2183	2632	gi 39978	P16 [Bacillus subtilis]	54	39
316	1	3	767	gi 143264	membrane-associated protein [Bacillus subtilis]	54	34
318	7	5035	4565	gi 606080	ORF_o290; Geneplot suggests frameshift linking to o267, not found Escherichia coli	54	28
319	3	1393	2163	gi 148327	vancomycin response regulator [Enterococcus faecium]	54	34
323	2	1256	2560	gi 413940	ipa-16d gene product [Bacillus subtilis]	54	26
364	7	7335	7724	gnl PID e250171	F18C12.1 [Caenorhabditis elegans]	54	31
386	5	2399	3844	gi 155369	PTS enzyme-II fructose [Xanthomonas campestris]	54	37
392	3	2004	3353	gi 872306	integral membrane protein [Streptomyces pristinaespiralis] pir S57509 S57509 integral membrane protein - Streptomyces ristinaespiralis	54	32
424	5	1553	1371	gi 160316	major merozoite surface antigen [Plasmodium falciparum] sp P50495 MSPI_PLAPP MEROZOITE SURFACE PROTEIN 1 PRECURSOR MEROZOITE SURFACE ANTIGENS (PMMSA) (GP195)	54	37
445	2	1897	1178	gi 1781503	MigA [Pseudomonas aeruginosa]	54	31
452	5	2506	2805	gi 216292	neopullulanase [Bacillus sp.]	54	34
457	2	2178	1024	gi 405570	Trak protein shares sequence similarity with a family of proteins ncoded on Gram-negative gene transfer systems such as TraD from the plasmid [Plasmid pSK41]	54	35
461	3	627	1418	gi 797332	MocD [Agrobacterium tumefaciens]	54	38
466	5	5419	3770	gi 1652892	ABC transporter [Synechocystis sp.]	54	29

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
475	3	2745	1990	gi 532546	ORF13 [Enterococcus faecalis]	54	35
495	1	2	295	gi 304990	ORF_o290 [Escherichia coli]	54	21
502	4	3518	3216	gi 1573270	hemolysin (tlyC) [Haemophilus influenzae]	54	33
510	5	3089	3931	gi 1732200	PTS permease for mannose subunit IIPMan [Vibrio furnissii]	54	29
570	1	1	930	gi 1001582	penicillin-binding protein 1A [Synecocystis sp.]	54	31
573	6	2763	3164	gi 416197	homologous to plasmid R100 pemK gene [Escherichia coli]	54	35
590	1	433	2	gi 5323309	25 kDa protein [Escherichia coli]	54	33
643	2	1202	1477	gnl PID e125689	256 kD golgin [Homo sapiens]	54	29
705	1	2	682	gi 148921	LicD protein [Haemophilus influenzae]	54	39
730	1	370	167	gnl PID e245531	ORF YLR068w [Saccharomyces cerevisiae]	54	29
745	1	502	209	gi 581140	NADH dehydrogenase [Escherichia coli]	54	37
749	1	413	3	gi 664840	TagB [Dictyostelium discoideum]	54	44
932	1	3	320	gi 537207	ORF_f277 [Escherichia coli]	54	27
4	6	5671	4748	gi 216267	ORF2 [Bacillus megaterium]	53	34
16	8	6231	6806	gi 517105	spermidine acetyltransferase [Escherichia coli]	53	35
17	1	2	2497	gi 387880	collagen adhesin [Staphylococcus aureus]	53	35
42	4	2942	3529	gi 1633572	Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus]	53	20
69	6	3149	4879	gi 1486244	unknown [Bacillus subtilis]	53	30
72	3	1455	2063	gi 1592197	M. jannaschii predicted coding region MJ1576 [Methanococcus jannaschii]	53	32
79	1	83	592	gi 633757	pr2 [Mycoplasma hyopneumoniae]	53	28
83	8	5179	4412	gi 496100	unknown function; putative [Bacteriophage phi-LC3]	53	39
85	10	7180	6764	gi 1303940	YgiU [Bacillus subtilis]	53	35
92	2	789	986	gi 1372996	Rho [Borrelia burgdorferi]	53	28

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
95	10	7546	7734	gi 162379	variant surface glycoprotein [Trypanosoma brucei]	53	28
99	4	1391	1861	gi 1499620	M. jannaschii predicted coding region MJ0798 [Methanococcus jannaschii]	53	34
100	44	29982	29749	gi 1590997	M. jannaschii predicted coding region MJ0272 [Methanococcus jannaschii]	53	35
102	5	4787	5089	gi 1399011	immunogenic secreted protein precursor [Streptococcus pyogenes]	53	40
113	1	825	4	gnl PID e264148	unknown [Mycobacterium tuberculosis]	53	24
114	4	6555	5113	gi 487282	Na <sup>+</sup> -ATPase subunit J [Enterococcus hirae]	53	33
119	6	3581	3994	gi 473707	positive regulator for virulence factors [Clostridium perfringens]	53	31
123	19	16463	18115	gi 1591361	NADH oxidase [Methanococcus jannaschii]	53	33
136	1	381	4	gi 152744	IpaD protein [Shigella flexneri]	53	32
138	9	8079	7594	gi 467371	LACI family of transcriptional repressors (probable) [Bacillus subtilis]	53	29
142	8	4594	4007	gi 755216	N-acetylmuramidase [Lactococcus lactis]	53	38
162	12	12482	11937	gi 1063250	low homology to P20 protein of Bacillus licheniformis and bleomycin acetyltransferase of Streptomyces verticillius [Bacillus subtilis]	53	36
163	1	546	31	gi 153767	ORF [Streptococcus pneumoniae]	53	34
163	7	4973	3453	gi 29468	beta-myosin heavy chain (1151 AA) [Homo sapiens]	53	36
167	2	1038	2006	gi 413930	ipa-6d gene product [Bacillus subtilis]	53	27
173	11	8865	7843	gi 1778569	YaaF homolog [Escherichia coli]	53	39
190	8	6842	3549	gi 387880	collagen adhesin [Staphylococcus aureus]	53	38
199	2	2725	950	gi 1652570	nitrate transport protein NrtB [Synechocystis sp.]	53	32
200	13	6184	5954	gi 1652679	hypothetical protein [Synechocystis sp.]	53	40
200	17	9287	7890	gi 1574246	H. influenzae predicted coding region	53	35

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
205	6	2048	3229	gi 148026	H11409 [Haemophilus influenzae]		
211	2	270	1052	gi 483940	topoisomerase III [Escherichia coli]	53	32
221	10	5119	5994	gi 1353529	transcription regulator [Bacillus subtilis]	53	30
232	7	4344	3925	gi 1665759	ORF12 [Bacteriophage rlt]	53	44
238	21	18705	19247	gi 1574062	Similar to Schistosoma mansoni amino acid permease (L25058). [Homo sapiens]	53	35
239	1	2	1636	gi 433932	hypothetical [Haemophilus influenzae]	53	30
250	1	1469	318	gi 987094	activator of (R)-hydroxyglutaryl-CoA dehydratase [Acidaminococcus fermentans]	53	35
253	4	1759	1028	gi 537245	membrane transport protein [Streptomyces hygroscopicus]	53	22
271	8	4649	5800	gi 413966	aspartokinase I-homoserine dehydrogenase I [Escherichia coli] pir S56629 S56629	53	35
276	26	15786	15112	gi 1699017	aspartate kinase (EC 2.7.2.4) / homoserine dehydrogenase (EC 1.1.1.3) - Escherichia coli	53	27
279	11	8309	7797	gi 1651934	ipa-42d gene product [Bacillus subtilis]	53	26
288	8	3997	4872	gi 43943	ErpB2 [Borrelia burgdorferi]	53	26
290	6	4391	5680	gi 466882	hypothetical protein [Synecocystis sp.]	53	35
294	3	1197	1481	gi 173004	second subunit of EII-Sor [Klebsiella pneumoniae]	53	32
330	3	2351	3367	gi 466691	pps1; B1496_C2_189 [Mycobacterium leprae]	53	29
334	8	8172	9182	gi 1652483	topoisomerase I [Saccharomyces cerevisiae]	53	40
368	1	620	102	gi 487273	No definition line found [Escherichia coli]	53	34
377	4	2424	2260	gi 221407	hypothetical protein [Synecocystis sp.]	53	29
382	1	257	36	gi 1592016	Na+ -ATPase subunit I [Enterococcus hirae]	53	29
					FP5 [Fowlpox virus]	53	35
					M. jannaschii predicted coding region MJ1371 [Methanococcus jannaschii]	53	32

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
387	1	2	460	gi 1574317	repressor protein (GP:L22692_1) [Haemophilus influenzae]	53	30
394	10	8379	10412	gi 882463	protein-N(pi)-phosphohistidine-sugar phosphotransferase [Escherichia coli]	53	34
399	4	2349	3098	gi 453287	OmpR protein [Escherichia coli]	53	27
420	2	1378	719	gi 1437473	nitrate transporter [Bacillus subtilis]	53	28
441	6	5361	7937	gi 1592205	M. jannaschii predicted coding region MJ1595 [Methanococcus jannaschii]	53	38
461	1	6	512	gi 1651800	L-glutamine:D-fructose-6-P amidotransferase [Synechocystis sp.]	53	29
497	3	1700	1960	gi 4328	RIF1 gene product [Saccharomyces cerevisiae]	53	33
503	1	669	4	gnl PID e202290	unknown [Lactobacillus sake]	53	30
538	2	1053	262	gi 1613769	response regulator [Streptococcus pneumoniae]	53	30
539	6	6172	5183	gi 567887	putative repressor [Streptomyces peucetius]	53	32
551	1	629	162	gi 1256649	putative [Bacillus subtilis]	53	26
557	1	9	695	gi 143177	putative [Bacillus subtilis]	53	31
569	2	418	1158	gi 1184684	MucD [Pseudomonas aeruginosa]	53	26
614	1	99	581	gi 485280	28.2 kDa protein [Streptococcus pneumoniae]	53	32
660	1	1	279	gnl PID e288480	R10E8.f [Caenorhabditis elegans]	53	34
776	1	3	635	gi 151352	mandelate racemase (EC 5.1.2.2) [Pseudomonas putida]	53	33
11	2	1117	1656	gi 143150	levR [Bacillus subtilis]	52	29
17	6	5327	6559	gnl PID e250887	potential coding region [Clostridium difficile]	52	37
19	31	17760	17978	gi 1079556	dShc [Drosophila melanogaster]	52	42
19	38	20306	22627	gnl PID e139448	host interacting protein [Bacteriophage B1]	52	32

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
25	4	2662	2087	gi 1072067	PepF [Rhodobacter sphaeroides]	52	23
25	6	5596	3407	gi 1303866	Yqgs [Bacillus subtilis]	52	34
49	3	1135	1569	gi 496279	putative [Bacteriophage Tuc2009]	52	25
53	1	850	2	sp P32697 YBHE_ECO LI	HYPOTHETICAL 30.2 KD PROTEIN IN MODC 3' REGION.	52	35
54	9	10909	2687	gi 1633572	Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus]	52	30
57	6	4779	8402	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	52	31
58	6	6446	5949	gnl PID e255921	F53F4.10 [Caenorhabditis elegans]	52	31
72	13	13446	13195	gi 532541	ORF8 [Enterococcus faecalis]	52	37
81	17	13692	12520	gi 1732203	GLcNAc 6-P deacetylase [Vibrio furnissii]	52	35
84	1	3	1355	gi 64288	fast skeletal muscle Ca-ATPase [Rana esculenta]	52	34
100	2	1917	1027	gi 1353560	ORF43 [Bacteriophage r1t]	52	34
101	1	30	1862	gi 405957	YeeF [Escherichia coli]	52	24
106	8	8517	7600	gi 454904	rfbG gene product [Shigella flexneri]	52	41
108	1	1	1059	gnl PID e255337	unknown [Mycobacterium tuberculosis]	52	29
123	4	2899	3495	gi 1305720	prs-associated putative membrane protein [Escherichia coli]	52	24
128	23	17561	16740	gi 473805	'regulatory protein sfsI involved in maltose metabolism' [Escherichia coli]	52	32
130	8	6693	7481	gi 1552775	ATP-binding protein [Escherichia coli]	52	30
138	1	40	1359	gi 1045867	oligoendopeptidase F [Mycoplasma genitalium]	52	31
138	2	2757	1384	gi 1591425	hypothetical protein (GP:X91006_2) [Methanococcus jannaschii]	52	26
138	6	6317	5940	gi 1486247	unknown [Bacillus subtilis]	52	36
142	10	7337	5466	gi 1151158	repeat organellar protein [Plasmodium chabaudi]	52	34
149	1	33	1133	gi 1762962	FemA [Staphylococcus simulans]	52	31

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
161	1	3	245	gi 151276	histidine utilization genes repressor protein (hut) [Pseudomonas utida]	52	35
163	4	2048	1320	gi 1064810	function unknown [Bacillus subtilis]	52	27
164	8	4882	5103	gi 57251	precursor (AA -35 to 1766) [Rattus norvegicus]	52	38
165	9	7247	7474	gi 1652671	hypothetical protein [Synechocystis sp.]	52	28
178	5	1887	1681	gi 220704	CAMP-dependent protein kinase catalytic subunit-beta [Rattus sp.] gi 191177 CAMP-dependent protein kinase beta-catalytic subunit Cricetulus sp.]	52	36
180	24	22536	23774	gi 581052	cytosine deaminase [Escherichia coli]	52	28
190	9	8891	7056	gi 1592079	M. jannaschii predicted coding region MJ1429 [Methanococcus jannaschii]	52	39
195	8	2000	2272	gi 868024	HIC-1 gene product [Homo sapiens]	52	52
202	11	9189	10145	gi 141861	traA gene product [Plasmid pAD1]	52	33
204	4	1361	2011	gi 1184118	mevalonate kinase [Methanobacterium thermoautotrophicum]	52	33
204	8	4018	5142	gnl PID e283860	carotenoid biosynthetic gene ERWCRTS homolog [Sulfolobus solfataricus]	52	31
208	2	1112	2296	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	52	35
215	1	772	2	gi 1480429	putative transcriptional regulator [Bacillus stearothermophilus]	52	26
218	4	4072	3425	gi 862630	glyceraldehyde-3-phosphate dehydrogenase [Buchnera aphidicola] sp Q07234 G3P_BUCAP GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC .2.1.12) (GAPDH)	52	35
228	1	1	741	gnl PID e264148	unknown [Mycobacterium tuberculosis]	52	29
230	2	149	634	gi 437705	hyaluronidase [Streptococcus pneumoniae]	52	28
233	8	6166	4982	gi 1001708	Nifs [Synechocystis sp.]	52	31
240	3	725	967	gi 399655	Ca2+ regulatory protein [Saccharomyces	52	21

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
288	7	3171	4028	gi 147403	cerevisiae] sp P35206 CSG2_YEAST CSG2 PROTEIN PRECURSOR. mannose permease subunit II-P-Man [Escherichia coli]	52	27
318	1	7	819	gi 1303849	YggB [Bacillus subtilis]	52	33
330	1	1062	154	gi 144859	ORF B [Clostridium perfringens]	52	29
330	9	6815	7213	gi 1439527	EIIA-man [Lactobacillus curvatus]	52	31
345	9	8348	9397	gi 606292	ORF_0696 [Escherichia coli]	52	27
398	3	2671	1877	gi 144859	ORF B [Clostridium perfringens]	52	29
411	1	992	3	gnl PID e283950	daunorubicin resistance ATP-binding protein DrrA [Sulfolobus solfataricus]	52	27
422	2	1292	585	gi 537214	yjg gene product [Escherichia coli]	52	32
436	2	1669	1205	gi 507323	ORF1 [Bacillus stearothermophilus]	52	29
450	1	119	754	gi 1573916	multidrug resistance protein (emrB) [Haemophilus influenzae]	52	32
453	1	190	381	gi 182021	elastin [Homo sapiens]	52	40
455	7	5767	4634	gnl PID e155312	integrase [Bacteriophage TP901-1]	52	34
479	1	138	758	gi 1742859	ORF_ID:0327#7; similar to [SwissProt Accession Number P54449] [Escherichia coli]	52	27
517	1	763	2	gi 152780	rhamnosyl transferase II [Shigella dysenteriae]	52	29
518	3	1735	848	gi 153858	wall-associated protein [Streptococcus mutans]	52	20
526	3	2297	1848	gi 147402	mannose permease subunit III-Man [Escherichia coli]	52	27
617	1	1	462	gi 142863	replication initiation protein [Bacillus subtilis]	52	35
639	3	1068	259	gi 1591153	hypothetical protein (sp:P46348) [Methanococcus jannaschii]	52	30
703	1	773	81	gi 793910	surface antigen [Homo sapiens]	52	31



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
737	1	235	2	gi 666000	hypothetical protein [Bacillus subtilis]	52	29
791	4	1368	1802	gnl PID e269549	Unknown [Bacillus subtilis]	52	28
825	1	1	300	gi 732538	No definition line found [Caenorhabditis elegans]	52	28
981	1	226	2	gi 951100	P45016a-msl [Mus spretus]	52	36
17	23	23542	22163	gi 1652483	hypothetical protein [Synechocystis sp.]	51	32
65	6	4302	3691	gi 397498	Membrane Ribose Binding Protein [Bacillus subtilis] pir S42714 S42714 membrane ribose-binding protein - Bacillus ubtilis	51	31
69	5	2926	2537	gi 1773150	hypothetical 14.8kd protein [Escherichia coli]	51	30
92	1	973	44	gnl PID e243523	ORF YGR130c [Saccharomyces cerevisiae]	51	29
103	6	5272	3593	gi 312940	threonine kinase [Streptococcus equisimilis]	51	32
111	7	4195	3317	pir G64143 G64143	hypothetical protein HI0143 - Haemophilus influenzae (strain Rd KW20)	51	29
115	7	4526	3414	gi 405879	yeiH [Escherichia coli]	51	27
123	29	27788	28207	gi 147402	mannose permease subunit III-Man [Escherichia coli]	51	27
125	1	223	2	gi 4482	SLV1 gene product [Saccharomyces cerevisiae]	51	37
128	21	16156	15638	gi 606026	ORF_ol15 [Escherichia coli]	51	27
137	4	3207	5369	gi 1673692	(AE000005) Mycoplasma pneumoniae, C09_orf422 Protein [Mycoplasma pneumoniae]	51	26
138	28	18295	18771	gi 149647	ORF2 [Listeria monocytogenes]	51	31
145	6	4054	5271	gi 1653860	N-acyl-L-amino acid amidohydrolase [Synechocystis sp.]	51	41
155	4	3019	2273	gi 1486242	unknown [Bacillus subtilis]	51	41
180	8	7951	9189	gi 1657522	hypothetical protein [Escherichia coli]	51	32
186	2	859	1620	gi 511497	oleoyl-acyl carrier protein thioesterase [Coriandrum sativum]	51	29

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
186	3	1644	2060	sp P37348 YECE_ECO LI	HYPOTHETICAL PROTEIN IN ASPS 5'REGION (FRAGMENT).	51	38
194	3	1521	1276	gi 332697	fusion protein [Human parainfluenza virus 2]	51	32
195	7	1986	3767	gi 405570	Trak protein shares sequence similarity with a family of proteins ncoded on Gram-negative gene transfer systems such as TraD from the plasmid [Plasmid pSK41]	51	28
197	1	3	494	gi 1592234	DNA topoisomerase I [Methanococcus jannaschii]	51	32
198	2	1521	862	gi 1196483	unknown protein [Lactobacillus casei]	51	32
238	16	13630	14730	gi 1772652	2-keto-3-deoxygluconate kinase [Haloferax alicantei]	51	36
257	5	5646	4513	pir S43367 S43367	metallothionein - Green crab, common shore crab	51	38
261	6	4950	4519	gi 581545	orf 4 [Staphylococcus aureus]	51	26
270	5	4480	4220	gi 1066975	F49E2.5a [Caenorhabditis elegans]	51	28
306	10	5928	6905	gi 1752736	gene required for phosphorylation of oligosaccharides/ has high homology with YJR061w [Saccharomyces cerevisiae]	51	28
324	3	1590	2405	gi 409925	VirR positive regulator [Streptococcus pyogenes]	51	25
328	2	632	309	gi 466475	putative phospho-beta-glucosidase [Bacillus stearothermophilus] pir D49898 D49898 cellobiose phosphotransferase system celC - acillus stearothermophilus	51	30
340	2	898	1152	gi 40046	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] ir S15936 NUBSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	51	39

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
340	4	3617	2445	gi 763052	integrase [Bacteriophage T270]	51	33
379	10	11742	11311	gi 887829	D21141 uses 2nd start; frame determined by Lac fusion [Escherichia coli]	51	34
380	1	2	1123	gi 309662	pheromone binding protein [Plasmid pCF10]	51	34
395	1	526	95	gi 490986	phi 105 repressor orf2 [unidentified]	51	27
424	4	2512	995	gi 1633572	Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus]	51	31
444	1	737	483	gi 1245376	cardiac ryanodine receptor [Oryctolagus cuniculus]	51	34
483	1	1	642	gi 1303981	YqkD [Bacillus subtilis]	51	29
500	1	2	550	gi 987094	membrane transport protein [Streptomyces hygroscopicus]	51	23
525	3	492	983	pir A57438 A57438	tryptophan-rich sensory protein - Rhodobacter sphaeroides (strain 2.4.1)	51	38
534	1	2	1165	gi 147516	ribokinase [Escherichia coli]	51	33
547	1	1	387	gi 1353528	ORF11 [Bacteriophage rlt]	51	33
553	2	1728	1330	pir B55124 B55124	thioredoxin - Chlorobium sp.	51	27
574	1	2291	2476	bbs 129435	RprX=inner membrane signal-transducing protein [Bacteroides fragilis, Peptide, 519 aa] [Bacteroides fragilis]	51	36
574	2	3145	3420	gi 1732202	PTS permease for mannose subunit IIMan N terminal domain [Vibrio furnissii]	51	29
594	2	530	225	gi 1657696	tryptophan hydroxylase [Gallus gallus]	51	40
605	3	1220	1936	gnl PID e289149	similar to B. subtilis YcsE hypothetical protein [Bacillus subtilis]	51	32
609	1	1027	74	gi 1226279	strong similarity to Schistosoma amino acid permease (GB:L25068) [Caenorhabditis elegans]	51	26
656	2	2033	2950	gi 143213	putative [Bacillus subtilis]	51	26
670	1	1508	369	gi 1652222	hypothetical protein [Synechocystis sp.]	51	25

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
673	1	2	1135	gi 532553	ORP20 [Enterococcus faecalis]	51	27
674	2	1158	778	gi 467451	unknown [Bacillus subtilis]	51	26
735	2	477	725	gi 757791	aromatic amino acid permease (Corynebacterium glutamicum) pir S52754 S52754 aromatic amino acid permease - Corynebacterium glutamicum	51	38
924	1	794	3	gi 40663	sialidase [Clostridium septicum]	51	35
4	5	3811	4728	gi 413948	ipa-24d gene product [Bacillus subtilis]	50	29
8	3	3310	2180	gi 1592205	M. jannaschii predicted coding region MJ1595 [Methanococcus jannaschii]	50	28
11	9	5269	5520	gi 1651800	L-glutamine:D-fructose-6-P amidotransferase [Synechocystis sp.]	50	25
12	6	9045	8662	gnl PID e254943	unknown [Mycobacterium tuberculosis]	50	23
15	4	2911	4269	gi 1592173	N-ethylamine chlorohydrolase [Methanococcus jannaschii]	50	28
19	10	4934	5530	gi 825569	unknown [Saccharomyces cerevisiae]	50	20
28	5	7515	7057	gi 1230586	orf10; Method: conceptual translation supplied by author [Vibrio cholerae O139]	50	38
45	9	4279	5019	gi 1591029	thioredoxin/glutaredoxin [Methanococcus jannaschii]	50	32
54	16	7739	7590	gi 1589837	cuticle preprocollagen [Meloidogyne incognita]	50	46
59	5	1551	2345	gi 144297	acetyl esterase (XynC) [Caldocellum saccharolyticum] pir B37202 B37202 acylesterase (EC 3.1.1.6) (XynC) - Caldocellum saccharolyticum	50	34
62	3	1650	1360	gnl PID e205266	LEA76 homologue type2 [Arabidopsis thaliana]	50	31
91	10	8858	7521	gi 758229	integrase [Bacteriophage phi-13]	50	31
112	5	3548	2133	gi 1184262	GadC [Shigella flexneri]	50	25
123	13	13099	14319	gi 178273	alanine:glyoxylate aminotransferase [Homo sapiens]	50	31

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
123	15	14395	15675	gi 467342	unknown [Bacillus subtilis]	50	28
123	31	28700	29494	gi 43942	first subunit of EII-Sor [Klebsiella pneumoniae]	50	27
124	2	1666	1061	gi 556016	similar to plant water stress proteins; ORF2 [Bacillus subtilis] gi 556016 similar to plant water stress proteins; ORF2 [Bacillus subtilis]	50	34
128	39	32767	31829	gi 39993	UDP-N-acetylmuramoylalanine--D-glutamate ligase [Bacillus subtilis]	50	33
135	11	8803	7694	gi 895747	putative cel operon regulator [Bacillus subtilis]	50	26
138	21	14648	13653	gi 1591472	malic acid transport protein [Methanococcus jannaschii]	50	26
146	3	2338	1415	gi 1732200	PTS permease for mannose subunit IIPMan [Vibrio furnissii]	50	27
160	2	724	1302	gnl PID e264218	F54F3.4 [Caenorhabditis elegans]	50	30
164	15	15432	16364	gi 409286	bmrU [Bacillus subtilis]	50	27
167	9	17082	15394	gi 143156	membrane bound protein [Bacillus subtilis]	50	30
179	3	2350	4485	gi 1408485	yxDM gene product [Bacillus subtilis]	50	24
180	30	31056	30643	gnl PID e254644	membrane protein [Streptococcus pneumoniae]	50	27
184	1	2	1015	gi 834232	cymE gene product [Klebsiella oxytoca]	50	24
194	7	4335	4817	gi 1256652	25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]	50	30
195	29	11712	12422	gi 662263	ORF5 [Plasmid pIP501]	50	25
204	1	2	166	gi 328656	envelope polypeptide [Human immunodeficiency virus type 1]	50	45
205	7	3118	3861	gi 437697	traE [Plasmid RP4]	50	31
216	11	7181	7750	gnl PID e254644	membrane protein [Streptococcus pneumoniae]	50	30
223	10	7036	8082	gi 606423	T09B9.1 [Caenorhabditis elegans]	50	30

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
223	22	19257	19799	gi 1256141	YbL [Bacillus subtilis]	50	29
233	4	3102	2320	gi 887826	GUG start [Escherichia coli]	50	32
238	6	5102	3906	gi 1161219	homologous to D-amino acid dehydrogenase enzyme [Pseudomonas aeruginosa]	50	29
239	3	4449	5159	gi 41519	P30 protein (AA 1-240) [Escherichia coli]	50	31
242	2	147	2210	gi 160299	glutamic acid-rich protein [Plasmodium falciparum] pir A54514 A54514 glutamic acid-rich protein precursor - Plasmodium alciparum	50	30
248	2	263	712	gi 143725	putative [Bacillus subtilis]	50	32
256	8	8531	7395	gnl PID e250452	C44H9.4 [Caenorhabditis elegans]	50	38
265	3	1150	893	gi 1402527	ORF6 [Enterococcus faecalis]	50	39
276	24	14203	14000	gi 1591019	M. jannaschii predicted coding region MJ0297 [Methanococcus jannaschii]	50	33
276	32	20601	19924	gi 1334905	BXLF2 late reading frame, encodes gp85; homologous to RF 37 VZV and glycoprotein H of HSV (gPIII of VZV) [Human herpesvirus 4]	50	29
286	1	1	747	gnl PID e257895	homology with truncated ORF2 of pepF2 [Lactococcus lactis]	50	32
301	17	11706	13313	gi 562039	NADH dehydrogenase, subunit 2 [Acanthamoeba castellanii] pir S53835 S53835 NADH dehydrogenase chain 2 - Acanthamoeba astellanii mitochondrion (SGC6)	50	26
338	5	2206	3729	gi 829194	bacterial cell wall hydrolase [Enterococcus faecalis] pir A38109 A38109 autolysin - Enterococcus faecalis sp P37710 ALYS_ENTFA AUTOLYSIN (EC 3.5.1.28) N-ACETYL-MURAMOYL-L-ALANINE AMIDASE)	50	34
345	12	11781	13379	gnl PID e235181	unknown [Mycobacterium tuberculosis]	50	32

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
360	2	2879	408	gi 40782	bps2 gene product [Desulfohalobus ambivalens]	50	25
372	1	6	440	gi 1552733	similar to voltage-gated chloride channel protein [Escherichia coli]	50	31
372	2	391	738	gi 1591749	TRK system potassium uptake protein A [Methanococcus jannaschii]	50	23
377	3	2262	1846	gi 52797	kinesin heavy chain [Mus musculus]	50	22
392	1	433	2	gi 147213	phnP protein [Escherichia coli]	50	33
399	31	29803	30186	gi 146288	PTS enzyme III glucitol [Escherichia coli]	50	30
518	4	2885	2040	gi 475107	regulatory protein [Pediococcus pentosaceus]	50	29
528	1	3	665	gi 215098	excisionase [Bacteriophage 154a]	50	38
562	1	631	107	gi 1592205	M. jannaschii predicted coding region MJ1595 [Methanococcus jannaschii]	50	28
596	1	227	1153	gi 963039	orf gene product [Enterococcus hirae]	50	26
680	1	2	1090	gi 1050297	product p150Glued [Neurospora crassa]	50	27
755	1	2	430	gi 1736469	Tetracenomycin C resistance and export protein. [Escherichia coli]	50	33
838	1	428	3	gi 530424	50S ribosomal protein [Mycoplasma capricolum]	50	30
14	2	3453	538	gi 47049	asal gene product (AA 1-1296) [Enterococcus faecalis] ir S10223 HMSO1F aggregation protein asal - Enterococcus faecalis asmid pAD1	49	25
56	7	5367	4822	gi 924754	glycine reductase complex selenoprotein B [Clostridium litorale]	49	31
68	9	4741	7389	gi 1591494	M. jannaschii predicted coding region MJ0797 [Methanococcus jannaschii]	49	21
94	10	9425	6633	gi 1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis]	49	30
98	12	12306	11701	gi 1303784	YqeD [Bacillus subtilis]	49	26

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
117	7	4789	6228	gi 435493	orf4 gene product [Lactococcus lactis]	49	26
123	21	18576	19745	gi 298032	EF [Streptococcus suis]	49	29
125	4	2358	1594	gnl PID e237295	unknown [Saccharomyces cerevisiae]	49	27
125	6	4235	3453	gi 1573885	glycosyl transferase (lgtD) [Haemophilus influenzae]	49	32
144	5	3715	4062	gi 507130	emm64 gene product [Streptococcus pyogenes]	49	30
162	8	10472	9120	gi 47045	NADH oxidase [Enterococcus faecalis]	49	34
179	18	18426	17848	gi 40060	DNA polymerase III (AA 1-1437) [Bacillus subtilis] p P13267 DP3A_BACSU DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7).	49	27
180	19	18727	19917	gi 143000	proton glutamate symport protein [Bacillus stearothermophilus] pir S26247 S26247 glutamate/aspartate transport protein - Bacillus stearothermophilus	49	31
224	1	145	1371	gi 1103862	TolA [Pseudomonas aeruginosa]	49	32
236	8	10955	9249	gi 431272	lysis protein [Bacillus subtilis]	49	28
278	1	757	2	gi 467478	unknown [Bacillus subtilis]	49	29
290	8	6860	7366	gi 466875	nifU; B1496_C1_157 [Mycobacterium leprae]	49	35
318	5	4065	3190	gi 144859	ORF B [Clostridium perfringens]	49	25
318	8	6052	5033	gi 1439528	FIIC-man [Lactobacillus curvatus]	49	30
335	1	534	40	gi 216861	24K membrane protein [Pseudomonas aeruginosa]	49	24
338	4	2861	2169	gnl PID e288536	F37H8.a [Caenorhabditis elegans]	49	30
346	4	1257	2273	gi 536970	ORF_f543 [Escherichia coli]	49	25
355	20	12902	15262	gi 292836	trichohyalin [Homo sapiens]	49	20
366	1	1	1437	gi 405857	yehU [Escherichia coli]	49	26
375	8	7663	6470	gi 1573546	H. influenzae predicted coding region HI0561 [Haemophilus influenzae]	49	30
377	2	1624	392	gi 532553	ORF20 [Enterococcus faecalis]	49	27
399	5	3960	3142	gi 1742362	nta operon transcriptional regulator.	49	29



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
456	1	1070	342	gi 290533	[ <i>Escherichia coli</i> ] similar to <i>E. coli</i> ORF adjacent to suc operon; similar to gntr class f regulatory proteins [ <i>Escherichia coli</i> ]	49	27
619	1	2	232	gi 665956	ribosomal protein S20 homolog [ <i>Aeromonas sobria</i> ] sp P45786 RS20_AERHY 30S RIBOSOMAL PROTEIN S20 (FRAGMENT). sp P45788 RS20_AER50 30S RIBOSOMAL PROTEIN S20 (FRAGMENT).	49	41
621	1	319	942	gi 149456	nisin-resistance protein [ <i>Lactococcus lactis</i> ]	49	29
630	1	3	1190	gi 537145	ORF_f437 [ <i>Escherichia coli</i> ]	49	34
736	1	859	2	gi 1592020	hypothetical protein (SP:P37555) [ <i>Methanococcus jannaschii</i> ]	49	27
849	1	232	11	gi 145514	cyclopropane fatty acid synthase [ <i>Escherichia coli</i> ]	49	35
47	11	14140	13307	gi 1045937	M. genitalium predicted coding region MG246 [ <i>Mycoplasma genitalium</i> ]	48	34
103	4	2492	1605	gi 1591514	membrane protein [ <i>Methanococcus jannaschii</i> ]	48	19
127	7	6836	5736	gi 1573128	hypothetical [ <i>Haemophilus influenzae</i> ]	48	24
138	22	14742	15590	gi 580884	ipa-89d gene product [ <i>Bacillus subtilis</i> ]	48	33
160	6	3048	3665	gi 1652295	serine esterase [ <i>Synechocystis</i> sp.]	48	28
162	3	3048	2491	gi 143830	xpaC [ <i>Bacillus subtilis</i> ]	48	13
193	2	1257	310	gi 1591153	hypothetical protein (SP:P46348) [ <i>Methanococcus jannaschii</i> ]	48	24
219	1	61	573	gnl PID e257628	ORF [ <i>Lactococcus lactis</i> ]	48	32
221	11	5952	6428	gi 1303733	YqaN [ <i>Bacillus subtilis</i> ]	48	31
232	4	2776	1712	gi 142707	comG2 gene product [ <i>Bacillus subtilis</i> ]	48	24
236	6	8618	7689	gi 550075	cephalosporin-C deacetylase [ <i>Bacillus subtilis</i> ]	48	26
238	28	25896	26825	gi 47906	rha regulatory protein [ <i>Salmonella</i> ]	48	31

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
251	2	1935	640	gi 1143026	typhimurium		
252	1	2036	3	gnl PID e228699	ORF10 [Spitoplasmavirus] homologous to yqbo of the skin element [Bacillus subtilis]	48	30
269	1	481	2	gi 1045975	sensory rhodopsin II transducer [Mycoplasmagenitalium]	48	37
315	5	4604	2649	gi 396400	similar to eukaryotic Na <sup>+</sup> /H <sup>+</sup> exchangers [Escherichia coli] sp P32703 YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS INTERGENIC REGION (O549).	48	28
327	1	128	916	gi 216314	esterase [Bacillus stearothermophilus]	48	30
330	6	4486	5337	gi 43942	first subunit of EII-Sor [Klebsiella pneumoniae]	48	21
330	7	5325	6230	gi 147404	mannose permease subunit II-M-Man [Escherichia coli]	48	33
345	10	9571	10521	gi 1736789	Collagenase precursor (EC 3.4.-.-). [Escherichia coli]	48	26
509	1	1	444	gi 606376	ORF_ol62 [Escherichia coli]	48	33
531	1	624	109	sp P50848 YPWA_BAC SU	HYPOTHETICAL 58.2 KD PROTEIN IN KDGT-XPT INTERGENIC REGION.	48	33
549	3	962	369	gi 1001212	molybdenum cofactor biosynthesis protein C [Synechocystis sp.]	48	32
725	1	3	500	gi 1151158	repeat organellar protein [Plasmodium chabaudi]	48	25
789	1	133	717	gi 42724	rhas (AA 1-278) [Escherichia coli]	48	39
936	1	32	316	gi 532549	ORF16 [Enterococcus faecalis]	48	45
2	2	2662	449	gi 929878	J1027 gene product [Saccharomyces cerevisiae]	47	20
4	2	1002	2192	gi 763052	integrase [Bacteriophage T270]	47	29
21	8	6350	5355	gi 1066343	mu-crystallin [Homo sapiens]	47	29
25	3	915	2048	gi 1064813	homologous to sp:PHOR_BACSU [Bacillus subtilis]	47	21

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
59	2	953	1378	gi 872306	integral membrane protein [Streptomyces pristinaespiralis] pir S57509 S57509 integral membrane protein - Streptomyces pristinaespiralis	47	26
81	7	4970	4206	gi 1591754	hypothetical protein (SP:P39364) [Methanococcus jannaschii]	47	22
82	3	1534	866	gi 397526	clumping factor [Staphylococcus aureus]	47	21
110	5	2313	3767	gi 151928	48 kDa protein [Rhodobacter sphaeroides]	47	26
150	11	7839	9107	gnl PID e275490	C30H6.k [Caenorhabditis elegans]	47	16
161	2	116	1450	gnl PID e283830	aminotransferase [Sulfolobus solfataricus]	47	23
165	8	8081	6129	gi 924925	heparinase III protein [Cytophaga heparina]	47	29
180	31	31515	31054	gi 1591753	N-acetylglucosamine-1-phosphate transferase [Methanococcus jannaschii]	47	29
194	11	8247	9236	gi 1480429	putative transcriptional regulator [Bacillus stearothermophilus]	47	26
225	2	1039	701	gi 1212992	Protein sequence and annotation available soon via Swiss-Prot; available at present via e-mail from LABEIT@EMBL-Heidelberg.DE [Homo sapiens]	47	33
232	1	196	969	gi 293033	integrase [Bacteriophage phi-LC3]	47	30
232	6	3687	3340	gi 142706	comG1 gene product [Bacillus subtilis]	47	28
233	10	8424	6739	gi 887816	possible start 13 codons upstream, for o765 [Escherichia coli]	47	35
346	2	706	1083	gi 536970	ORF_f543 [Escherichia coli]	47	27
352	1	112	843	gi 1591857	H+-transporting ATPase [Methanococcus jannaschii]	47	28
410	1	3	980	gi 1652869	NADH dehydrogenase [Synechocystis sp.]	47	30
465	2	1976	1749	gi 211659	p68 protein; c-rel proto-oncogene [Gallus gallus]	47	30
491	3	3752	2466	gi 881434	ORFP [Bacillus subtilis]	47	24
501	1	48	809	gi 467429	unknown [Bacillus subtilis]	47	33

Table 3: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
532	1	3	287	gi 755724	alpha-toxin [Clostridium novyi]	47	32
578	1	707	81	gi 532547	ORF14 [Enterococcus faecalis]	47	30
605	4	2051	2470	gi 1783233	hypothetical [Bacillus subtilis]	47	22
626	3	2459	2169	gi 1573573	2',3'-cyclic-nucleotide 2'-phosphodiesterase (cpdB) [Haemophilus influenzae]	47	44
650	1	1042	341	gi 404802	integrase [Saccharopolyspora erythraea]	47	26
665	1	714	1175	gi 143655	sporulation protein [Bacillus subtilis]	47	22
754	2	1086	736	gi 143835	PBSX repressor [Bacillus subtilis]	47	27
845	1	2	241	gi 1303952	YqjA [Bacillus subtilis]	47	26
911	1	1	456	gi 1019640	ORFX (a homolog to the prgX gene of the pheromone response plasmid pCF10); putative [Plasmid pHKK701]	47	26
933	1	16	303	gi 331002	first methionine codon in the ECLF1 ORF [Salmirine herpesvirus 2] gi 60394 ORF 73; ECLF1 [Salmirine herpesvirus 2]	47	29
17	17	13073	13675	gi 1304597	abortive phage resistance protein [Lactococcus lactis]	46	27
19	11	5515	6393	gi 1353529	ORF12 [Bacteriophage rlt]	46	28
42	3	2460	3011	gi 1064814	homologous to sp:PHOP_BACSUB [Bacillus subtilis]	46	33
49	9	4042	5793	gnl PID e59644	predicted 86.4kd protein; 52Kd observed [Mycobacteriophage 15]	46	22
74	6	4039	3434	gi 143542	RNA polymerase sigma-30 factor [Bacillus licheniformis] pir B28625 SZBSSL transcription initiation factor sigma H - acillus licheniformis	46	27
89	14	14259	12967	gi 1499089	M. jannaschii predicted coding region MJ0305 [Methanococcus jannaschii]	46	32
89	15	15737	14427	gi 1653339	hypothetical protein [Synechocystis sp.]	46	22
94	13	12634	11132	gi 1402515	membrane-spanning transporter protein [Clostridium perfringens]	46	23

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
100	18	13493	11958	gi 15470	portal protein [Bacteriophage SPp1]	46	31
144	2	2364	1126	gnl PID e183450	hypothetical EcsB protein [Bacillus subtilis]	46	25
144	9	8977	6236	gi 710421	unknown [Staphylococcus aureus]	46	24
152	7	3397	4557	gnl PID e254991	hypothetical protein [Bacillus subtilis]	46	25
158	7	7144	5993	gi 1045800	ribose transport system permease protein [Mycoplasma genitalium]	46	28
180	11	10882	10055	gi 303953	esterase [Acinetobacter calcoaceticus]	46	23
181	3	1173	976	gi 1591638	M. jannaschii predicted coding region MJ0975 [Methanococcus jannaschii]	46	36
240	1	715	221	gi 1766062	Ats1 [Schizosaccharomyces pombe]	46	28
254	2	499	2	gi 153661	translational initiation factor IF2 [Enterococcus faecium] sp P18311 IF2_ENTFC INITIATION FACTOR IF-2.	46	32
262	4	5276	4431	pir A45605 A45605	mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum	46	20
309	1	2	673	gi 1651714	type 4 prepilin peptidase [Synechocystis sp.]	46	40
312	1	18	872	gi 580884	ipa-89d gene product [Bacillus subtilis]	46	32
324	6	4450	4836	gi 1061418	ArsC [Plasmid R46]	46	28
345	1	2241	1333	gi 144859	ORF B [Clostridium perfringens]	46	24
386	4	1438	2421	gi 405894	1-phosphofructokinase [Escherichia coli]	46	31
395	8	3584	3853	gnl PID e120267	sucrose-phosphate synthase [Beta vulgaris]	46	25
491	2	2527	1169	gnl PID e267595	Unknown, similar to peptidases [Bacillus subtilis]	46	29
495	3	612	869	gi 406286	triase phosphate/phosphate translocator [Flaveria pringlei] pir S37553 S37553 triase phosphate/3-phosphoglycerate/phosphate translocator - Flaveria pringlei	46	27
513	1	2	946	gi 143024	glucose-resistance amylase regulator	46	26

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
					[Bacillus subtilis] pir S15318 S15318 ccPA protein - Bacillus subtilis sp P25144 CCPA_BACSU GLUCOSE-RESISTANCE AMYLASE REGULATOR CATABOLITE CONTROL PROTEIN)		
520	3	914	2674	gi 1163086	microfilarial sheath protein SHP3 [Brugia malayi]	46	27
554	1	3	788	gi 413972	ipa-48r gene product [Bacillus subtilis]	46	27
568	1	1574	3	gi 532549	ORF16 [Enterococcus faecalis]	46	28
809	1	506	135	gi 49021	surface exclusion protein (SEAL) [Enterococcus faecalis] ir S22452 S22452 surface exclusion protein seal precursor - terococcus faecalis plasmid PAD1	46	28
813	1	2	1090	gi 150556	surface protein [Plasmid pCF10]	46	34
78	2	4915	2516	gi 577295	The hal225 gene product is related to human alpha-glucosidase. [Homo sapiens]	45	20
81	9	6123	5386	gi 147200	phnF protein [Escherichia coli]	45	28
85	1	120	761	gi 457514	glcC [Bacillus subtilis]	45	19
94	11	10681	9668	gi 289753	homology with nucleolin protein; putative [Caenorhabditis elegans] pir S44897 S44897 ZK1236.2 protein - Caenorhabditis elegans sp P34618 Y082_CAEEL HYPOTHETICAL 33.8 KD PROTEIN ZK1236.2 IN HROMOSOME III.	45	23
108	3	2427	1789	gnl PID e263931	OrfD [Streptococcus pneumoniae]	45	27
108	4	3338	2352	gi 606150	ORF_f309 [Escherichia coli]	45	25
131	6	3981	5309	gi 1590845	hypothetical protein (PIR:S51413) [Methanococcus jannaschii]	45	36
144	11	10215	8944	gi 1001554	hypothetical protein [Synecocystis sp.]	45	30
164	11	8247	6736	gi 409925	VirR positive regulator [Streptococcus pyogenes]	45	22
192	1	1598	591	gi 1736826	Lysozyme M1 precursor (EC 3.2.1.17) (1,4-b-N-acetylmuramidase M1). [Escherichia	45	27

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% Ident
					coli]		
223	16	14409	15212	gi 1651958	hypothetical protein [Synechocystis sp.]	45	32
279	7	5236	5772	gi 1736514	Isochorismatase (EC 3.3.2.1) (2,3 dihydro-2,3 dihydroxybenzoate synthase). [Escherichia coli]	45	29
364	3	2419	4098	gi 309662	pheromone binding protein [Plasmid pCF10]	45	26
459	1	2	307	gi 1679640	ORFA [Mycoplasma mycoides mycoides SC]	45	27
491	1	1022	135	sp P27434 YFGA_ECO LI	HYPOTHETICAL 36.2 KD PROTEIN IN NDK-GCPE INTERGENIC REGION.	45	20
496	1	847	2	gi 1208489	serum resistance locus BrkB [Synechocystis sp.]	45	19
542	2	1169	804	gi 1064811	function unknown [Bacillus subtilis]	45	28
63	3	1047	1919	gi 39848	U3 [Bacillus subtilis]	44	26
93	3	1108	1374	sp Q04747 SRF2_BAC SU	SURFACTIN SYNTHETASE SUBUNIT 2.	44	27
155	10	8354	7620	sp P35136 SERA_BAC SU	D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH).	44	29
215	2	2192	1134	gi 468760	ORF334 [Rhizobium meliloti]	44	31
303	1	466	2	gi 431950	similar to a B.subtilis gene (GB: BACHEMEHY.5) [Clostridium asteurianum]	44	22
310	1	284	39	pir S01294 S01294	intermediate filament protein B - Roman snail	44	26
311	1	122	2668	gi 532549	ORF16 [Enterococcus faecalis]	44	27
320	1	709	2	gi 290801	member of super-family of ABC proteins [Francisella tularensis (var. ovidica)]	44	23
341	14	13882	12998	gi 142863	replication initiation protein [Bacillus subtilis]	44	16
345	15	16445	18001	gi 151282	DL-hydantoinase [Pseudomonas sp.]	44	34
386	3	1340	570	sp P46117 YARA_PRO ST	HYPOTHETICAL 31.5 KD PROTEIN IN AARA 3'REGION.	44	19
862	1	483	4	gi 929796	precursor of the major merozoite surface	44	26

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
19	3	1695	1372	gi 603263	antigens [Plasmodium alciparum]		
45	17	14045	14995	gnl PID e233895	yel055p [Saccharomyces cerevisiae]	43	31
57	1	667	317	gi 664840	hypothetical protein [Bacillus subtilis]	43	32
71	2	1537	2568	gi 1303981	TagB [Dictyostelium discoideum]	43	22
72	18	20511	20164	gi 349045	YqkD [Bacillus subtilis]	43	26
94	9	6581	6039	gi 1146245	merozoite surface antigen 2 [Plasmodium falciparum]	43	36
180	17	16391	17656	gi 1290540	putative [Bacillus subtilis]	43	28
252	2	2407	1829	gi 154381	f445 [Escherichia coli]	43	24
276	30	19091	18480	gi 15470	chemoreceptor [Salmonella typhimurium]	43	19
311	2	2666	4639	gi 160299	portal protein [Bacteriophage SPP1]	43	23
631	2	1126	2328	gi 1519696	glutamic acid-rich protein [Plasmodium falciparum] pir A54514 A54514 glutamic acid-rich protein precursor - Plasmodium alciparum	43	28
11	3	1509	2342	gi 143150	coded for by C. elegans cDNA yk126f9.5;	42	21
45	14	10730	12028	gi 666069	coded for by C. elegans cDNA yk159h6.3;	42	23
72	19	21070	21981	gnl PID e236595	coded for by C. elegans cDNA yk126f9.3;	42	23
123	35	32205	32768	gi 1772652	coded for by C. elegans cDNA yk159h6.5	42	27
136	5	2737	2375	gi 153858	coded for by C. elegans cDNA yk159h6.5 [Caenorhabditis elegans]	42	27
167	4	2701	6540	gi 1519696	levR [Bacillus subtilis]	42	27
					orf2 gene product [Lactobacillus leichmannii]	42	23
					orf7 gene product [Enterococcus faecalis]	42	23
					2-keto-3-deoxygluconate kinase [Haloferax alicantel]	42	27
					wall-associated protein [Streptococcus mutans]	42	27
					coded for by C. elegans cDNA yk126f9.5;	42	27
					coded for by C. elegans cDNA yk159h6.3;	42	27
					coded for by C. elegans cDNA yk126f9.3;	42	27
					coded for by C. elegans cDNA yk159h6.5	42	27



Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
195	31	12430	13155	pir S33124 S33124	[Caenorhabditis elegans]		
211	1	187	2	gi 1653346	tpr protein - human	42	24
					GDP-mannose pyrophosphorylase	42	33
					[Synecocystis sp.]		
242	13	8089	12447	gi 951460	FIM-C.1 gene product [Xenopus laevis]	42	31
305	5	4354	5340	gi 1408485	yxm gene product [Bacillus subtilis]	42	25
355	18	9964	12549	gi 532549	ORF16 [Enterococcus faecalis]	42	30
446	4	4428	5261	gi 47528	glucosyltransferase S [Streptococcus salivarius]	42	25
656	3	2866	3456	gi 142857	MreD protein [Bacillus subtilis]	42	25
686	11	3646	3921	pir A44805 A44805	eggshell protein - fluke (Schistosoma haematobium) (subclone SH.E 2-1)	42	42
920	1	41	316	gi 532549	ORF16 [Enterococcus faecalis]	42	40
23	3	729	487	gi 414525	meiotin-1 [Lilium longiflorum]	41	41
56	5	3511	2324	gi 1591610	probable ATP-dependent helicase [Methanococcus jannaschii]	41	21
98	17	16843	16274	gi 1742129	Immunity repressor protein. [Escherichia coli]	41	23
167	6	6734	9811	gnl PID e249616	F56H9.1 [Caenorhabditis elegans]	41	37
171	13	10879	11871	gi 331002	first methionine codon in the ECLF1 ORF [Saimiriine herpesvirus 2] gi 60394 ORF 73; ECLF1 [Saimiriine herpesvirus 2]	41	23
181	2	1012	500	gi 455315	ORF 4 [Plasmid pIP404]	41	24
230	4	3664	3224	gi 498251	glutamate/aspartate transporter II [Homo sapiens]	41	22
718	1	2	613	gi 984656	ORF3 [Salmonella typhimurium]	41	22
19	30	16391	17770	gi 806704	Upf2p [Saccharomyces cerevisiae]	40	21
164	16	16440	17951	gi 348056	trans-acting positive regulator [Bacillus anthracis]	40	22
200	12	5956	4841	gi 1574243	H. influenzae predicted coding region HI1405 [Haemophilus influenzae]	40	24

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
216	10	6799	7194	gi 146279	glucitol-specific enzyme III (gutB) [Escherichia coli]	40	27
292	13	8633	10741	gi 1008233	ORF YJL076w [Saccharomyces cerevisiae]	40	18
345	13	14050	15333	gi 581051	cytosine permease [Escherichia coli]	40	25
521	1	177	1466	gi 289614	homology with glucose induced repressor, GRR1; putative Caenorhabditis elegans]	40	18
64	3	2646	1855	gi 154924	spectinomycin adenyltransferase [Transposon Tn554]	39	27
100	17	12037	10565	gi 1052806	product required for head morphogenesis [Bacteriophage SPPI]	39	24
529	1	326	4939	gi 295671	selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III [Saccharomyces cerevisiae]	39	19
49	2	518	931	gi 166162	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	38	19
54	19	11264	10854	gi 160186	circumsporozoite protein [Plasmodium vivax]	38	31
164	21	22793	23587	gi 603857	secreted acid phosphatase 2 (SAP2) [Leishmania mexicana]	38	18
167	3	2322	2756	gi 435039	proline-rich cell wall protein [Gossypium hirsutum]	38	36
204	2	133	798	gi 396401	No definition line found [Escherichia coli]	38	25
475	2	761	1792	gi 1574532	H. influenzae predicted coding region HI1680 [Haemophilus influenzae]	38	27
164	19	20738	21385	gi 165704	[Rabbit smooth muscle myosin light chain kinase mRNA, complete DS.], gene product [Oryctolagus cuniculus]	37	20
394	6	5649	6395	gi 603857	secreted acid phosphatase 2 (SAP2) [Leishmania mexicana]	36	16
958	1	1	459	gi 951460	FIM-C.1 gene product [Xenopus laevis]	36	28

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% Sim	% Ident
399	21	16383	21359	gi 1707247	partial CDS [Caenorhabditis elegans]	34	13
150	12	9056	11740	gi 1015903	ORF YJR151c [Saccharomyces cerevisiae]	33	19
195	34	13017	15512	gi 632549	NF-180 [Petromyzon marinus]	33	18

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
2	1	458	3
2	3	2208	2624
5	3	928	1440
8	6	4792	5877
8	7	5480	5262
12	1	2	832
12	2	771	4622
13	1	2	1684
14	1	531	130
15	2	862	1197
16	1	51	200
17	4	3309	3665
17	13	10079	10261
17	18	14431	13682
17	22	21525	21956
17	27	27055	27567
18	4	2172	1591
18	5	2524	2249
18	7	3467	3715
18	8	4082	3555
18	9	4333	4055
18	10	4395	4204
18	11	4498	4677
18	12	4656	5393
18	13	5878	5492
18	15	6296	6931
19	1	1047	676
19	2	1068	1247
19	4	1747	2031
19	5	2244	2612
19	7	2797	2943

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
19	9	3873	4730
19	13	6884	7420
19	14	7428	8042
19	16	9246	8425
19	17	9412	9615
19	19	9733	9918
19	20	10032	10334
19	21	10422	11009
19	22	11516	11944
19	24	12423	12881
19	26	14606	15427
19	27	15414	15848
19	28	15802	16134
19	29	16064	16393
19	32	17846	18052
19	33	18021	18356
19	34	18334	18684
19	35	18659	19036
19	36	18991	19677
19	37	19671	20132
19	39	22603	23337
19	40	23319	25580
21	2	762	262
21	5	3440	2925
21	10	7684	7241
23	5	2098	2652
23	8	4912	4709
23	9	4911	5246
23	10	5087	5353
23	22	14318	14926
23	23	14924	15565

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
23	24	15559	16083
23	29	17567	18022
25	2	553	1005
25	5	3363	2653
26	2	1220	1654
27	1	297	4
28	1	239	2833
29	5	3244	2822
29	6	4014	3301
29	7	4168	4557
29	8	5620	4595
32	3	2646	1375
32	4	2573	3010
39	9	4636	4986
40	2	1346	981
43	1	120	620
43	4	1972	2280
45	3	1557	1961
45	4	2012	2230
45	5	2218	2553
45	11	7226	5670
45	12	7270	10113
45	13	10013	10732
46	1	42	872
46	2	886	1125
46	4	2807	3100
47	4	5101	5625
47	10	13239	12847
49	1	106	504
49	8	2858	4132
49	10	5777	6193

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
49	11	6166	6720
52	5	3505	3110
52	7	5160	5603
52	8	5662	5459
54	2	400	729
54	4	1326	1610
54	5	2354	1335
54	6	1676	2080
54	7	2151	2576
54	12	4181	3954
54	13	5975	6289
54	14	6869	7144
54	15	7433	7107
54	18	9764	11086
55	2	252	440
56	2	1344	658
57	9	12450	12605
58	7	7066	6425
59	3	1350	952
59	4	1225	1515
59	7	2958	3200
62	6	4116	3007
63	1	77	364
63	2	455	1060
63	7	5422	5910
63	8	5870	6751
63	9	6688	7296
64	2	1849	1523
64	4	3183	2644
64	5	3422	3213
65	5	3787	3389

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
65	7	5043	4300
65	8	5354	4959
65	9	7005	6328
67	6	3719	4060
68	2	569	348
68	5	3234	2821
68	6	3808	3221
68	10	7495	8106
70	2	2102	1614
70	3	2019	2231
71	3	3362'	3787
72	21	22464	22709
72	22	22690	23019
72	23	23013	23834
73	1	154	2
74	1	61	486
74	3	1334	1981
75	4	3227	2136
75	5	3994	3251
75	6	3348	3632
75	7	4519	4043
75	8	4296	4529
75	10	6518	5769
76	2	1079	1897
76	4	2113	2436
76	6	4737	4105
77	3	1874	2704
77	4	2665	2459
78	3	5814	5398
79	3	848	1645
79	4	2121	1642



Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
81	8	5392	4961
81	13	8428	8874
81	21	15746	14802
82	1	858	4
82	2	198	383
83	3	2194	2604
83	4	2728	2405
83	6	2855	3172
83	10	7188	6184
83	11	7415	7065
83	17	12259	12561
83	21	15890	16456
83	23	16946	17251
84	5	7071	7949
85	7	6518	6174
89	2	1012	599
89	3	1382	939
89	4	2350	1370
89	5	2523	2314
89	9	7505	7182
89	16	15846	15673
89	19	20070	19045
90	1	3	689
91	7	3834	4127
91	8	4288	5268
91	9	7259	5748
91	12	9737	8973
91	13	10162	9731
92	3	1458	958
92	4	1934	1287
93	2	479	949

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
93	4	1344	1727
94	1	770	45
94	3	1460	1618
94	5	2279	1734
94	12	11000	10641
95	11	7674	7907
95	12	8604	8056
95	13	8725	8546
96	1	758	1018
96	2	1038	1469
98	5	6809	5994
98	10	10338	10652
98	11	10650	11558
99	2	232	513
100	4	3728	4048
100	6	5866	5378
100	7	6574	5921
100	8	6923	6534
100	9	7355	6921
100	10	7698	7339
100	11	8226	7744
100	13	9395	8514
100	15	10368	10102
100	19	14770	13505
100	20	15300	14758
100	21	15783	15298
100	23	17699	17292
100	25	20933	20625
100	26	21200	20946
100	28	23713	23156
100	29	23948	23691

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
100	30	24312	23965
100	31	24550	24287
100	32	24912	24565
100	33	25173	24910
100	34	26339	25158
100	36	27251	26994
100	37	27945	27232
100	39	28442	28227
100	40	28657	28403
100	46	30439	31146
100	47	31158	31712
101	2	850	464
101	3	2453	1899
102	6	5023	5616
102	9	6704	7111
103	7	5454	5296
105	2	1244	1828
106	4	5114	3294
106	6	7622	6168
106	7	6577	6867
108	6	5192	4158
110	1	2	454
110	6	3689	4207
110	9	9374	8553
110	10	9903	9361
110	11	10175	9843
111	6	3118	3267
112	4	2170	1043
114	2	1347	1135
116	8	4782	5147
117	4	2437	2670

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
117	6	3876	4640
117	8	5643	5927
117	9	6195	6488
117	12	9655	9837
119	1	3	500
119	2	670	1158
119	4	2730	2284
121	3	2276	3670
123	14	14304	14555
123	16	15305	15147
123	24	21896	22663
123	34	31458	32207
125	3	1581	1300
125	7	4516	4346
126	2	85	312
127	2	1047	787
127	3	2006	1299
127	4	3432	1924
128	4	3094	2747
128	5	3466	3305
128	6	4625	3507
128	7	4726	4550
128	13	8947	8522
128	15	9325	9582
128	17	10126	10380
128	24	17649	18038
129	1	276	1769
130	7	6478	6702
130	11	9386	9769
133	7	6622	7380
135	2	2289	1153

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
135	3	3380	2271
135	5	3778	3930
135	6	5835	5137
135	7	6649	5852
135	8	7021	6647
135	9	7420	7034
136	2	963	379
136	3	2009	939
136	4	2344	1973
138	4	5051	3636
138	11	8499	8753
138	12	8682	8536
138	13	8923	9270
138	14	9333	9887
138	15	9628	10308
138	16	10422	10216
138	23	15980	15678
138	24	16437	16063
138	30	19388	19828
139	3	1068	1466
139	4	3338	1983
139	5	3769	3317
139	6	4114	3818
139	7	4838	4236
139	10	5639	5175
142	1	369	106
142	2	1005	367
142	3	2140	980
142	4	2504	2127
142	5	2821	2474
142	6	3294	2806

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
142	7	4000	3635
143	1	650	3
143	3	1090	173
143	4	1044	433
144	10	7570	8403
144	12	10727	10335
145	1	188	30
145	2	775	978
150	9	6876	7166
150	13	11538	11242
152	1	35	445
152	2	405	914
152	3	912	1430
152	4	1349	2212
152	5	2210	2896
152	6	2739	3368
152	8	4479	4694
152	11	6647	7321
154	7	4557	4195
155	3	1227	2180
155	12	8726	9022
156	3	3179	2664
158	11	10876	11220
160	1	545	3
162	1	228	1349
162	2	2513	1653
162	7	9163	7664
162	9	10619	10990
162	11	11891	11427
163	3	1043	1234
163	5	3217	2021

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
163	6	3455	3198
163	8	5611	4931
163	9	5969	5580
163	10	6144	5926
164	2	1100	1687
164	9	5729	5259
164	10	6778	5639
164	12	8277	8450
164	17	18224	18526
164	24	24751	24536
164	27	25764	26369
165	1	17	481
165	2	2213	1389
165	12	9871	9689
165	14	11416	10367
166	3	1250	1669
167	5	3774	3439
167	7	10479	14498
167	10	17476	18768
168	2	665	393
172	9	7018	6701
172	10	7097	7930
173	1	2	412
173	3	2341	2024
173	6	4234	5055
173	9	7882	7295
173	10	7413	7571
173	14	12308	11748
174	4	2350	3021
174	5	3082	3498
178	3	866	1105

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
179	8	8115	7816
179	17	17407	17135
180	4	3524	4537
180	5	4686	5687
180	6	5897	6949
180	9	9721	9299
180	10	9996	9715
180	20	19805	19954
180	23	21808	21509
180	25	24127	26460
180	27	27977	27474
181	1	381	82
183	1	190	2
183	4	1849	2211
183	5	2350	2568
183	7	3592	2978
183	8	4176	3571
185	2	1260	1424
185	3	2722	1301
185	4	3612	2671
187	2	727	1302
187	3	1293	1745
187	5	2592	2173
189	1	18	2180
190	1	466	68
190	2	896	411
190	4	1878	2165
190	5	2740	2384
190	10	10281	8875
191	2	861	658
191	3	1096	827



Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
192	2	1881	1564
193	1	316	2
193	7	4667	3813
194	1	30	641
194	2	608	1582
195	1	2	433
195	2	431	943
195	3	1055	465
195	4	972	1487
195	5	1507	1995
195	6	3314	1851
195	9	3089	3529
195	10	3521	3312
195	12	6604	6837
195	13	7049	6786
195	14	6825	7700
195	15	7682	7047
195	16	7202	7417
195	18	8278	9036
195	20	8583	8837
195	21	8871	9602
195	22	9251	9403
195	23	9600	10022
195	25	10020	10226
195	26	11229	10024
195	27	10659	10946
195	28	10944	11318
195	30	12449	12246
195	32	13212	12505
195	33	12558	12773
195	35	13673	14011

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
195	36	14811	14143
195	38	16061	16363
195	39	16320	16799
195	40	16515	16333
196	1	608	1411
197	9	9269	9553
200	2	1103	249
200	3	1335	1033
200	4	1769	1284
200	5	2124	1747
200	6	2792	2106
200	7	3073	2708
200	8	3510	3061
200	9	4126	3467
200	10	4350	4042
200	11	4847	4368
200	14	6487	6182
200	15	6681	6499
200	18	10749	9307
200	20	11787	11464
200	22	12859	12410
201	1	509	105
201	3	3704	3237
202	7	5296	4817
205	2	117	323
205	5	1669	2148
206	2	546	196
206	3	841	632
206	4	1622	777
206	9	5466	5035
209	1	472	86

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
209	3	1510	1280
210	3	3175	2363
210	6	5281	4868
210	8	5619	6002
211	4	1708	3756
212	1	919	2
213	2	1107	1826
214	2	2106	1237
214	4	3677	3132
217	6	3548	3162
218	1	1	1218
218	3	2731	3378
218	5	4188	4667
219	3	1386	910
219	4	1595	1344
220	2	794	1144
221	1	110	295
221	2	326	880
221	4	1496	1825
221	5	1907	2200
221	6	2169	2555
221	8	3425	4246
221	9	4233	5111
221	12	6419	6757
221	13	6751	6987
221	14	6911	7120
221	16	7400	7909
221	17	7963	8199
221	19	8597	9079
222	17	11376	11597
223	6	5328	5008

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
223	12	12189	13307
223	13	13291	13716
223	14	13601	13434
223	17	15331	15068
223	19	15940	17160
223	21	17710	19089
223	23	19800	20708
223	25	22857	22027
223	26	22757	23365
225	1	756	394
225	5	3793	2945
226	1	141	536
226	2	521	871
228	8	5473	4835
229	7	6749	6057
232	2	1461	910
233	5	3359	3063
233	11	7226	7456
236	1	3	482
237	1	1	219
237	3	1197	991
237	5	2009	2329
237	6	2319	3056
237	8	3261	3701
237	10	3900	4763
237	11	4730	4963
238	11	9966	9238
238	19	16613	17728
238	29	26812	27663
239	2	1576	4245
239	5	6393	6956

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
239	6	6902	7237
240	5	1537	1809
241	1	228	1040
242	9	6581	7015
242	10	6988	7368
242	12	7488	7928
245	2	1670	1251
247	2	1558	1812
250	4	3210	2998
251	1	622	2
252	3	2598	2383
252	4	2911	2564
253	1	1	345
253	2	359	898
254	1	2	307
254	3	318	4
256	5	3768	4040
256	7	7292	6639
256	9	9589	8465
257	2	992	294
257	4	4528	3596
257	7	6894	6718
257	8	7252	6884
257	9	7986	7231
258	2	544	804
258	3	1224	2921
258	4	2964	2728
258	5	2919	3752
258	6	4120	5298
261	1	3	362
264	1	582	361

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
264	2	881	561
264	3	1367	879
264	4	1966	1361
264	5	2316	1945
264	6	2636	2295
264	7	3194	2634
264	8	3531	3055
265	2	398	817
265	4	1583	1071
265	6	3293	3009
265	7	3186	3046
266	1	451	2
266	4	1983	2225
266	7	2540	2325
268	1	798	1223
268	2	1912	1265
270	4	3977	4186
270	6	4397	4573
271	5	2719	3066
271	6	3041	3352
271	9	6278	5862
271	10	6550	5993
271	14	10291	10004
272	3	1870	1199
272	4	3378	1831
276	5	2350	1994
276	8	3702	3103
276	9	4441	3692
276	10	4595	4416
276	12	8173	7382
276	14	10001	9762

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
276	15	11065	9890
276	17	11642	11250
276	19	12892	12503
276	21	13302	13099
276	22	13663	13271
276	23	13995	13642
276	25	15065	14211
276	27	16293	15955
276	29	18482	16563
276	31	19951	19016
279	3	1469	1675
279	4	1600	1923
279	5	2269	2105
279	10	7698	7279
280	3	3138	2968
281	4	2055	2552
282	1	316	2
282	2	456	1232
282	3	1957	1346
283	1	1	450
283	3	1098	1556
283	5	2062	2238
283	7	3127	3312
286	3	2883	2698
287	4	2359	2180
290	10	8820	9074
290	11	9008	9172
291	2	1103	855
291	3	2622	1123
292	1	2	283
292	2	701	330

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
292	5	2459	2866
292	7	4252	4995
292	9	6704	7096
292	10	7066	7827
292	12	8377	8622
292	15	11502	12674
292	17	13326	13727
292	18	13738	14778
294	1	117	623
294	2	905	723
294	6	2496	2272
295	7	4274	4510
300	4	3525	3337
301	6	6714	4852
301	13	10150	9914
301	16	11316	11657
301	18	13199	14398
301	19	15724	14657
306	3	1135	2727
306	4	2742	4025
306	5	4004	4552
306	6	4527	5117
306	7	5131	5466
306	9	5642	5968
306	11	7000	8013
306	12	7926	8138
306	13	8180	8908
306	14	8899	9120
306	15	9118	9510
306	16	9508	9963
306	17	9964	11313



Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
306	18	11319	11570
306	19	11540	11707
306	20	11626	11856
310	2	1126	176
310	5	4215	3556
311	4	5671	6006
311	5	6173	6778
311	6	6833	7225
311	7	7236	7520
311	8	7492	7926
312	2	859	1506
312	3	1449	1808
312	4	2043	2306
313	4	3568	3122
319	1	3	881
319	2	832	1185
321	1	638	898
321	4	1862	2131
321	5	2168	2548
321	6	2470	3159
321	7	3069	3395
321	8	3461	3733
324	1	3	692
324	2	867	1592
324	4	2392	3021
327	6	5052	5213
330	5	3745	3464
333	2	998	717
333	3	947	1534
335	2	1024	521
338	11	8869	8591

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
340	5	3931	3608
341	6	3484	3155
341	7	4348	3482
341	8	6419	4332
341	10	9264	7672
341	11	10777	9245
341	12	12026	10779
343	1	459	262
343	4	3905	2661
345	4	3467	3201
345	14	15320	16447
345	16	18409	18927
345	18	19974	20465
347	1	763	1155
350	5	3273	2980
351	1	693	280
351	2	1268	654
351	3	1716	1222
353	4	2749	2546
354	1	2	298
355	16	8911	9399
355	19	12476	12904
355	22	15766	15608
355	23	17165	17461
355	25	18313	19104
355	26	19092	19598
355	27	19692	19495
355	28	19734	20198
355	29	20196	20471
356	2	2204	1536
356	4	2887	2537

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
356	5	3167	2859
357	1	381	4
360	3	3167	2877
361	1	7	909
363	1	1405	167
363	6	7178	8404
364	1	41	331
366	2	1386	1598
367	19	8690	8941
368	4	1786	1947
369	4	1652	1428
372	6	5262	4534
376	2	625	293
377	1	331	2
379	4	2975	3142
382	3	2951	3277
382	4	4183	3320
383	6	6158	5637
386	9	5725	6027
387	2	486	980
390	2	1668	2057
390	3	3499	2867
391	1	2	154
392	5	5163	5387
394	1	1	375
394	8	6437	7585
394	9	7542	7967
394	11	10354	10713
395	5	1957	2229
395	9	3869	4216
395	11	4571	4960

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
398	1	395	1180
399	7	5691	6134
399	10	7662	7820
399	14	10111	9845
399	22	16699	16481
399	29	28519	28244
401	1	189	4
401	2	178	1044
401	3	1038	2141
401	5	3517	3939
402	3	919	1269
404	1	578	12
405	1	293	643
405	3	1926	1501
407	1	80	406
407	4	3188	3670
408	5	3037	2681
408	6	3786	3475
410	2	811	1092
413	2	742	1314
413	3	1275	1532
414	2	908	678
414	3	1137	1889
414	4	2738	1959
416	3	1945	1709
418	1	3	350
418	2	331	930
419	2	619	296
419	4	937	773
419	5	1305	910
419	6	1183	1521

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
419	7	1859	1299
419	8	2170	1850
419	9	2483	2160
419	10	3399	2470
419	11	3708	3397
420	3	1649	1452
421	6	3983	3510
424	1	797	3
424	2	513	851
424	3	1029	733
424	6	1859	1551
424	7	3076	2780
425	1	52	384
425	2	1031	777
425	3	1127	1936
427	2	1488	1114
427	3	2114	1464
430	2	1334	1489
431	1	420	196
431	2	634	269
432	2	1133	1372
432	3	2014	1439
432	6	3869	3378
433	1	292	2007
435	1	706	131
435	2	1730	1047
439	1	1	627
441	1	1	513
441	7	10592	7974
443	1	31	744
447	2	744	322

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
449	1	3	212
449	2	471	286
449	3	551	393
451	1	823	314
452	2	322	714
452	6	2806	3342
452	7	3358	3792
454	1	1033	2
455	3	3214	3837
455	5	4078	4488
455	6	4965	4117
455	8	5123	5473
457	1	940	35
461	2	476	691
461	4	1548	1991
461	5	2322	1948
461	6	2664	2449
462	5	2810	2064
464	2	2162	1530
465	1	1762	38
465	3	2373	2050
467	2	652	1260
467	3	1149	1442
469	2	922	1101
470	2	971	1768
473	2	450	220
475	1	1	969
477	2	1064	843
482	1	1	534
484	1	130	543
484	2	1320	1159

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
487	2	1258	1929
488	2	509	162
488	4	2247	1945
489	1	1	396
489	2	560	255
490	2	1096	458
491	5	5167	4433
491	6	5975	5247
491	7	6811	6041
494	1	650	3
497	5	3351	3536
497	8	4757	4308
497	10	5229	5086
497	11	5967	5671
499	1	663	247
502	2	1324	851
504	1	3	650
507	2	727	906
507	3	840	1010
510	3	2056	2574
512	2	854	300
514	2	1067	669
518	5	3119	2970
520	1	3	467
520	2	452	231
520	4	2218	1859
521	2	988	821
522	1	409	885
524	1	579	4
525	1	1	144
525	2	86	352

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
529	2	5731	6147
533	1	1044	157
536	3	587	1462
539	7	6180	6662
540	1	198	476
543	3	2179	1835
543	4	2404	2177
543	7	3924	3700
544	2	1004	870
546	2	497	324
547	3	717	965
549	2	371	135
550	1	527	3
550	2	864	709
550	3	1540	1277
550	4	2039	1509
552	5	4681	5073
552	8	8390	8223
555	1	470	267
560	1	635	210
560	2	834	514
563	2	1215	1469
564	1	8	511
564	2	1019	555
564	3	577	744
565	1	321	4
565	5	1266	1619
567	2	1055	531
571	3	1149	886
573	1	208	666
573	2	651	1148



Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
573	5	2558	2809
575	1	262	2
584	1	268	110
584	4	1310	795
584	5	1329	1574
586	1	771	4
588	1	346	56
588	2	1078	434
589	1	1	555
591	1	217	2
592	2	674	868
593	1	190	2
593	3	1035	1268
601	1	77	274
601	2	172	576
602	2	759	415
604	6	2868	2416
606	1	271	798
607	2	633	797
613	1	420	82
616	2	593	435
616	4	975	730
619	3	641	817
620	1	863	3
621	2	1493	2014
627	1	113	763
628	1	2	163
631	1	1	516
631	3	1715	1521
633	1	280	2
634	3	1139	1387

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
637	2	1613	738
637	3	1597	2208
637	4	2242	2694
637	7	3550	4545
637	9	4767	5171
639	1	175	2
640	2	468	689
643	1	496	320
645	1	1	537
645	2	539	1024
647	1	64	855
647	2	1419	895
649	1	2	364
651	1	539	3
653	2	738	550
656	8	7784	8587
657	2	1356	967
657	3	1708	1376
661	1	2	244
664	3	1149	820
672	1	546	10
673	2	1207	1827
676	1	443	790
679	1	998	219
682	3	749	1171
685	1	176	511
685	2	498	199
685	3	480	947
685	4	1000	1443
686	4	1567	2001
686	5	3238	1712

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
686	7	2965	3435
686	8	3441	3067
686	9	3752	3339
686	10	3530	3826
688	2	628	894
689	2	582	331
690	1	275	90
690	2	487	248
696	1	239	9
696	2	1237	233
696	3	1424	1200
697	1	20	520
698	1	29	313
698	2	217	483
701	5	1061	1534
707	2	855	538
709	1	1	675
710	1	3	416
712	1	674	96
713	1	933	139
713	2	1125	1436
716	2	1226	765
721	1	3	371
726	1	543	94
729	1	19	210
731	1	532	2
736	2	309	644
738	1	561	4
740	1	488	3
749	2	20	475
751	1	1	456

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
751	2	454	774
753	1	76	729
754	1	761	21
755	2	345	539
756	1	1	375
764	2	528	1088
772	1	1	558
772	2	432	866
775	1	706	2
778	2	992	834
780	1	52	351
782	1	3	557
783	1	28	609
791	1	1	582
791	2	859	641
791	3	1235	711
797	1	2	289
797	2	287	3
801	2	598	191
805	1	1	414
806	1	392	3
810	1	3	317
810	2	407	3
815	2	443	282
819	1	39	668
830	1	291	4
830	2	476	162
834	1	561	46
834	2	953	453
837	1	3	317
837	2	320	589

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
839	1	1	753
841	1	1	489
855	1	308	3
861	1	1	330
863	1	451	221
870	1	21	503
890	2	1548	1255
895	1	3	140
896	1	2	400
897	2	244	498
902	1	1	300
904	1	294	4
910	1	143	3
917	1	36	518
918	1	3	167
918	2	116	373
920	2	243	515
922	1	669	259
926	1	2	394
927	1	119	556
928	1	493	179
930	1	526	344
933	2	257	418
936	2	243	683
937	1	341	3
942	1	58	228
945	1	318	4
953	1	254	48
959	1	1198	164
959	2	1740	1123
963	2	462	232

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
965	1	403	2
969	1	360	4
970	3	673	314
972	1	3	470
973	1	2	700
974	1	2	235
974	3	270	467
981	2	154	405
984	3	164	337

**SEQUENCE LISTING PLACE INDICATOR**

PAGES 280 TO 2076, WHICH ARE THE COMPLETE SEQUENCE  
LISTINGS FOR THIS APPLICATION ARE LOCATED AFTER THE  
DESCRIPTION, CLAIMS, ABSTRACT & DRAWINGS.

## (i) GENERAL INFORMATION:

(i) APPLICANT: Charles Kunsch  
Patrick J. Dillon  
Steven C. Barash

(ii) TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and  
Polypeptides

(iii) NUMBER OF SEQUENCES: 982

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.  
(B) STREET: 9410 Key West Avenue  
(C) CITY: Rockville  
(D) STATE: Maryland  
(E) COUNTRY: USA  
(F) ZIP: 20850

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
(B) COMPUTER: HP Vectra 486/33  
(C) OPERATING SYSTEM: MSDOS version 6.2  
(D) SOFTWARE: ASCII Text

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE: herewith  
(C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: A. Anders Brookes



281

(B) REGISTRATION NUMBER: 36,373

(C) REFERENCE/DOCKET NUMBER: PB369PCT

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8512

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4315 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

AAGTCAATCA CTTGCAAGTC GTTTTCTGTC ATATGGCCGA CTAGCTCAGT GACACCAGGA	60
ATATAGCCAC CAGGGAAAAT ATAACGATTA ATCCAAGCAT TTTTAGCCCC ACCTTGTTGG	120
CGACTGATCC CATGAATCAA CGCCGTACCT TTAGGCGCTA AATTCCGCTG AACGACATCA	180
AAATATTCAT GTAGATTTTC CGCACCGACA TGTTCAAACA TCCCAACACT CGTAATATGG	240
TCAAAGACT CTCCTTTTAA ATCACGATAA TCCATCAATT TGACAGTCAT TCGATCTTGT	300
AGATCTTCTT TTTCTATAAT ATGGCGAATA TGATGAAATT GCTCTTCACT TAATGTAATC	360
CCAGTTGCTT TGGCTCCATA TTCTTTCACC GCAGTTAAAA TTAACGTGCC CCAGCCGCAG	420
CCAATATCCA GTAAAGTGTC GCCCTCTTTG ATAAACAATT TATCTAAAAT ATGATGAACT	480
TTATTCACCT GCGCTTGTTT TAATGTATCT TCAGGCGTTT TAAAATAAGC ACATGAATAC	540
GTCATTGTTT GGTCAAGCCA TTTTTTGTA AAATCATTTT CTAGATCGTA ATGGCTGTGA	600
ATATCCTCTT GCGAACGTTT TTTTGAATGA CTTTCTTTAG GAAGCCATTT AATAAATTTA	660
GCATTGTGTA AAAAGCTATC CTTTTGGTTA TACACATCAT AAATCAGsGC TTGGATATCG	720
CCTTCGATTT CAATTTTGCG ATCCATGTAG GCTTCCCCTA AAGTTAACGA AGCGTTATTC	780
AGTAAATCCT TCACAGGAAT TTTTTCATTG AATACAATTT TAAAAACCGG ATCCCCCGAC	840
CCTTGCCCAT ACTCTTTGAC GGTACCATCC CAGTATGTGA CTTGTGTCTT TTTGAAAAA	900
GACCATTTAA ACAGTTGACT GTACGTTTCT TTTTCTAACA TTGCATTCCC TCCATTAAAT	960
ACCATTTGAA GCCAAAACAA AAAGAAGTCG CTTTCCGGTA GTTCGTCAAA ACAAACACCA	1020
CAGTCCGTTT TAAACTGAAG CACAGAAAAG TTATCACCCC TTCTATGTTT CGCTTCTTTT	1080
TTTGCAATTA CAGTTCTATT CTACTCCTCT TTTAAAAATT TGAACATTCT TTTAACGTAA	1140
TACCTACTAT TGTTATTCTT TATCACAAAA AACTAGAGC CAGTCCTTGA CAGACTCCTC	1200
TAGTTCTAAA TATTATGCTT TCTTACGCAT CCGTTGTTCC GCATGAGTGT AAGCGCCATG	1260

CCATACGTGC CCCACATAAG GATTAACCTTG AATACCGTGT TTAATCGCCG CTGCTACAAA	1320
TTTTTTTGCT AAAGTTACTG CTTCTAACAC CGAATAACCT TTCGCCAAGC CAGCTGTGAT	1380
TGCCGCTGAA AAAGTACAAC CTGCACCATG ATTATAATCA GTAGGATATA ATTCAATTTTC	1440
CAAAAGATGC GCGGTGTGAC CATCGTAAAA CAAGTCCAGT GCTTTTTCAC CAGCTAAGCG	1500
ATGTCCCCCT TTAACCACGA CATGCTTGGC TCCCATTGT ACAATTCGTT TTGCCGCTTC	1560
TTCCATCTCC GCCACGGAAG AAATTTCGCC TAAACCAGAT AAGATGCCCG CTTCAATTAA	1620
ATTAGGCGTG GCCACTAATG CTAATGGCAG TAAATCGTTT TTTAGGCCTT CCACACTTTT	1680
GGGTTCGAGA ATTTGTGCCG TTCCCTTACA AGCAATGACT GGGTCAATCA CGACTTTTTG	1740
AATTTTCTCT TGTTTAATGT ACTTACTAGC CATTTTAATA TTTTGTCAT TACCCATCAT	1800
CCCTGTTTTT AAAGCCGCTA CTGGACCGCC TGCAAAAACC GAAATCAATT GTTTTTCTAA	1860
GAGCGTTTCT GGCAATTCAG TTAATTCATG TGACCAACCT GTCGTAGGAT CCATCGTCAC	1920
AATCGAGGTT AAAGTTGAAA ATCCAAAAAC TCCATACTCT TCAAATGTTT TTAAATCTGC	1980
TTGAATCCCT GCCCTCCAG TTGAATCGGA GCCTGCAATC GTCAATACTT TTTCCATTAA	2040
ATCACCTAAC CTTTTTCTCC AAGTATACTG AAGAAACAAG TCTGCTAAAA CAGCCAATTG	2100
GCTTATTTTT TAGCCAGCCA ATTTCTAAAC AAAAAAAGA CCAGAGAATA AATTCTCTGG	2160
TCTTACGTCC GAATACCCCA GTTTTTTACG CTGGTTAAAA CTATAGTTAA AAAGTTAATT	2220
ATTTAACGAT TTCAGTAACA ACGCCTGAAC CTACAGTACG TCCGCCTTCA CGAATAGAGA	2280
AACGAGTTCC GTCTTCGATA GCGATTGGGT GAATTAATTC AACGTCCATA GCAACGTTAT	2340
CACCAGGCAT TACCATTTC A GTACCTTCTG GCAATTCTAC AACACCAGTA ACGTCTGTTG	2400
TACGGAAGTA GAATTGAGGA CGGTAGTTAG TGAAGAATGG AGTGTGACGT CCGCCTTCTT	2460
CTTTTGATAA TACGTATACT TCAGCTTtGA ATTTgGTGTG TGGAGTGATT GTAGCTGGTT	2520
TAGCTAATAC TyGyCCACGT TCGATATCTT CACGTGCTAC ACCACGTatA AAGCACCGAT	2580
GTTGTCGCCT GCTTCAGCGT AGTCTAATAA TTTACGGAAC ATTTCAACAC CTGTAACAGT	2640
TGTTTTAGAT GTTTCGTCTT TAATACCAAC GATTTCAACT TCGTCACCAA CGCGAACTTC	2700
ACCACGTTCA ACACGGCCTG TAGCAACAGT ACCACGTCCA GTGATTGAGA ATACgTCTTC	2760
GACTGGCATC ATGAATGGTT TGTCAGTATC ACGTTCTGGA GTTGGGATAT ATTCGTCAAC	2820
TGCAGCCATT AATTCTAAGA TTTTTCCTTC ATAAGACTCG TCGCCTTCTA AAGCTTTCAA	2880
AGCAGAACCT GCGATAACTG GAACATCATC GCCTGGGAAA TCGTATTCTG ATAATAAGTC	2940
ACGAACTTCC ATTTCTACTA ATTCTAATAA TTCTTCGTCA TCAACCATAT CCATTTTGTT	3000
TAAGAATACA ACGATGTATG GTACACCAAC GTTACGTGAT AATAAGATAT GTTCACGTGT	3060
TTGAGGCATA GGACCATCAG CAGCAGAAAC TACTAAGATA GCTCCGTCCA TTTGAGCAGC	3120
ACCAGTGATC ATGTTTTTAA CGTAGTCCGC ATGTCCTGGG CAGTCaACGT GTGCaTAGTG	3180
ACGAGTTwCA GTTTCATATT CGATATGAGA AGTGTTGATT GTGATTCCAC GTTCTTTTTT	3240

TTCTGGAGCG TTATCGATAG AATCGTAGCT TTGTGCTTCC CCGCCACCGT GTTTTGATAA	3300
TACAGTAGCA ATTGCAGCTG TTAATGTAGT TTTACCATGG TCAACGTGTC CGATAGTACC	3360
AATGTTAACA TGGGATTTAG AACGGTCAAA TTTTCTTTT GCCATTTTGA ATGTTCTCTC	3420
TAAAAATATA AATTTTATTG TTGTTTTAT ATGTGATAGA CGTTACTAAA AATCGTCTGA	3480
TCATCATAAG CTATTTTACA CGAAATCTTG AAAAAAGGAT AGTTCTAAAA ACAAACTTT	3540
TATTCTCAAG GAAAATTATG CGTtGCCGCC GTTTTCTTG ATGATTCTT CTTGAACAGA	3600
CTTAGGTACG TCTTCATAGT GGTCAAATAC CATCATAAAT GTACCACGTC CTTGTGTGC	3660
TGAACGTAAT GTTGTAGCAT AACCAAACAT TTCAGCTAAA GGTACCATCG CATTACGAT	3720
TTGAGAGTTA CCGTGTGCTT CCATACCTTC AACACGTCCA CGACGACTTG TTACGTGTCC	3780
CATGATATCA CCTAAGTAAT CTTCAGGAAC AGTAATTGTA ACTTTCATCA TTGGTTCTAA	3840
AATAACTGGG TTTGCATTTT TAGCCGCAGC TTTCAATGCC ATAGAAGCAG CTACACGGAA	3900
GGCCGTTTCA TTTGAATCGA CATCGTGGTA TGAACCATCG TAAAGTTTGT CTTTGATATC	3960
AACTAATGGA TAACCAGCAA GAACACCGTT GTTCATAGAG TCTTCTAACC CTTTTCAAC	4020
TGCTGGGaTG TATTCACGAG GAACCACACC ACCGACAATT GCGTTTTCAA ACTCGAAGCC	4080
TTTACCTTCT TCATTTGGTG TAAATTCAAC CCaTAcGTGA CCGTATTGTC CTTTACCACC	4140
AGACTGACGA ACGAATTTAC CTTCTGCTTT AGTCGCAGCA CGGAATGTTT CGCGGTAAGA	4200
TACTTGAGGA GCACCTACGT TAGCTTCAAC TTTGAATTCA CGTTTCATAC GGTCTACTAA	4260
AACGTCTAAG TGCAATTCAC CCATaCCAGA GATAACTGTT TCACCAGTTT CAACC	4315

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2716 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ATTGTGCCAT AGATTGACC ATCTGCTGCT GTTGTAGTTT CCCAGTTTGT AATTTGGTAA	60
TGTTGTAAGA nCTTAGTTGG GACATTAATG GCAAAAGTTT TTTGGCGATA GGTATTAAAA	120
TTGGTTTTAC CCACCAACC TTCATAGGCT TCAAATTCAG CTCCACCTAA CAAAGGGTTG	180
TCGGAAATAA TCCAAGTGTC TTTTGTATGA TTCCGAATAA AACGAACGAT TGTACTGTTA	240
AAGCCTTTCG TTGTTGTACC GCCGTAATCT TTTTCTGCAG ACCAATGGTT CCAAACCGAG	300
TCATACTCTA AAGCATTCGG AAATTCATTG GTTGTAACC AACCGAGTGA ATTAATTTCT	360
TTACTTAGTT GACGACTTGC CCAACCTGAT TCGCCTTGGT TGCCCCAAAC ATCAACATAA	420
ATATAATCTA GATTCCGTGc TTTTtGTTTT AATcnTTAAA GCGCTCATAG CGACGACCAC	480
TCAATGTAtC GGGtCTTkGT TTAATAAAAT AAGACGGATC TAACCAATCC CAGCCACGCT	540

TCGTTGGATC AACTAATTCC TCATTAAATC CTTTTGCTTC TGGGTAAGAC TCGGTGTCAT	600
TAATATGCAC ACCGAAAACG GCATTTAATT TATGTCCTTC ATCAATTAAT TGATTCAACG	660
CTTGTTCCCC ACCAGGTCGC TGACCAATAG CACCGTAATC TGGATGCGCA GAATCATGTC	720
CTTCATTTTG ATACCCTTTT AGTAAATTCA TTTGTCCTAA TCCATCTGTT AAATTGTAAA	780
TACGTTTTGA TTCGTCTAAC GTCAC TAAGA ATGGGTTTGT CGCCTGACTA GCAAAGTTAA	840
AAGGAATCCG TTGGTTGACT AATTCAGGGA CTTTTTCCGC ACCCATTGGG TTATTCATAA	900
TTGAACGGAA CCCAATTGCC GCATCTTGCC AATCCACCCG ATTGTCGTCG TTGCTATCTT	960
TTGAGAATTT AACTTTCACT TCTGGCGTTT CTCCAGTTGT GTAATCTTCC GGTGCATCAA	1020
ATGGTCGATA GGTCCATGCC CCACTTGACA AGGTTGTTAC AAAACCAGTC GCCGCTTCTT	1080
TTGTTTGT TT ATGGATTCTGA TTGTGTGAC TACCATCGAC AGAGCCGTCT CCATAGGCGT	1140
TCGTCCAAAA ACTTGCAGCA TAATTCGCCG TATTTAAGAA ACCATACATA TATTTTTTCG	1200
TTTCTTCTTG GCTTGTTGTA GTTAAATCTT GGATAGTATC ACCATTTTTA TTTCCGTAT	1260
TTGCATTTGT CCCAGTATTC ATCACAACGC CAGCAAAAGT CGCCCCCTCA TCTGTACTAT	1320
TGACAGAAAT CAAGCCTTGA TTTGGAATTG AAATTGTGTG TATTTTGGTT CCTTCTTCAC	1380
GAATATCTGA GAGTTTCATA GCTAATTCTT GGCCTTCGGA AACAGTCAAT GTCGTTGTGA	1440
AACGCAAACC AATTTCTGGG ACAGCTACGT TATAAACGGC TTGaTTTGCT TCaCTCTTCT	1500
GTTTTTcAGC AGTTGCTTTA tATGCTTTGC CAyCAAyCAT TACTTGGTCa ATTTTCTCgT	1560
GTTGGCCATT AAAAACTACT TGATTACTCT GAGGGTCTTC GTAGCGAATA ATTCGTGGAA	1620
ATTGTGGATC AACACGGCT GTTAAATCTG CGGAAGAGAT TTTGTCGCCT TTTTCAGCTT	1680
CATTGTCTTT AGCAACAACA ATGCTGGCTT TTGCTTTGAT GTTTGTATTT TCCAAAGTAC	1740
CGTCGACTTC AAAAGTTCCT GGTCTTTAT ATGCATCAGG ATCGATTTC TCCAACGGA	1800
CTGCTTCATT GGCTCTTTA CCAGTATTAT AGGTCACCTT TACTTGTTC GGtAATTTTG	1860
GTGCGACACC AATTTTTGTA AAACTTTGA CTGGCGCAAT TTGTGTCACT TCTGGCACAA	1920
TTTCTGGAAT GCTATCTACG GGGCCATTTT TAAAATAGTC ATAATTGACA ATTTTTTTGT	1980
CGTACCACAA ACGGACACCC ACATGTCCGG CTTCTGTTGG AATTTTGTCT CCATTGGCCA	2040
AAACAGGTTT TCCTTCGTAA ATCAACGTGG TGTTAAGCCA GATAGTAATT TTTTCACCAA	2100
CATAACGTAC TTTCAACAAA TAATTCGTAT CTTCAATTTAA CGTTGGCCCA GAAATTGAAT	2160
CATTCCACGA ATTAGGGCTT TCAACTAACC ATTTGCCATT CGCGTTATAG CCTACAAAAA	2220
CCCAGCTATC TTTGGTATTA CCACGTATAA TCACACCCGT TCGACCGCTT CCAGCAGTGT	2280
ATTAAACTT TGTCTTACT TCACCATCAG CTAATTTAGG GGAATCTAAA TTAAAGATA	2340
CGGCATTATT ACCTGCTGCT GCATCTCGTG AAATCGCCAA TGA CTCTCCT GCTAATTCTC	2400
GATTTGTTTT ACCAATCACA TCTTGCCACT CACCAAGTTT CCCTCCTGGA AAATCAGATT	2460
GCCACAATTT TTCTTCTGTT GCGGTAGGTT CCACCATTGA TGAGCTTGTT TCAGTGCTGC	2520

TGTTTGTGT TTCTTCTGCA GACGTACTTA ATGGTACTAA AATCGTTGCG CTTGCCAATA	2580
GTGTCAATGT ACTAAATCGT TTTATTTTTC CATGTTTCAT CTGaTAAATC AGCCCCCTGA	2640
ACATTTTGG AACACGCTTT CATTTATTTA ACAATTAAAG GGGGGTTnCC AAAAATTAAA	2700
TTGTnTTCAT TTTTGG	2716

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1311 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TTGAAATTGC TAGCGTTTTTC CGCCATTGGC TATGTTTAGT CGCAACGGAG GGACCsCtGA	60
CTTTTTAGCG AAGCATcgTT GCCTGTAAAG aAAGTGACGA AAATTTCTAA TCCTGAAGGG	120
GAAACCGTGC TGGATGTCAT TCGCAATGGG AATGCGCAAG TGGTAATCaG TACCATGGAT	180
AAAAATCGTT CCAGTGcCAA CCAAGATGGA TTTTCAATCC GTCGGGAAGC AGTCGAACAT	240
GGGATTCCAT TGTTTACCTC ACTTGACACA GCAAATGCGA TTTTAAAAGT TTTAGAATCA	300
AGAGCCTTTA CGACAGAAGC CATTTAAGCA AAAATCTGGA ACAATCGTCT GAGGAGGACT	360
GTTCCAGATT CTTTTACGAA GAAATAGAAA ATCATTTATA GGGGAATCAG AGAATAAAGG	420
ACTGAAGGAT ACTGATGCAG AGAAAGCAAG AAATGATGAC CATTGTAGCT CAAAAACAGT	480
TAGCGCCACG GATTTATCAG TTAGACTTAC AAGGCGAGTT AGTGAAAGAG ATGACACGAC	540
CTGGACAGTT TGTGCATATC AAGGTTCCGC GTGCAGATTT ACTTTTGCGG CGACCGATTA	600
GCATCAACCA AATTGATCAT TCGAACGAAA CCTGTCGTTT GATTTATCGA GTCGAAGGCG	660
CTGGAACGGA AGTGTGTCG ACCATGAAGG CTGGTGAACA ATTAGATATT TTAGGTCCTT	720
TAGGAAACGG CTTTGATATA ACTACCGTAG CAGCTGGTCA AACTGCTTTT ATCGTTGGCG	780
GAGGAATCGG GATTCCCCCA TTGTATGAAT TATCAAAACA ACTCAATGAA AAAGGCGTGA	840
AAGTGATTCA TTTTCTTGGT TATGCATCAA AAGAGGTGTC TTATTACCAA CAAGAATTTA	900
TGGCATTAGG AGAAACACAT TTTGCCACAG ATGACGGCTC GTTTGGCGCT CATGGCAACG	960
TGGGCCGTTT ATTGTCAGAA GCGTTAGCAA AAGGACGGAT CCCTGATGCA GTGTATGCTT	1020
GTGGTGCGAA TGGTATGTTA AAAGCAATTG ATTCTTTATT TCCAACACAT CCACACGTTT	1080
ACCTTTCTTT AGAAGAACGG ATGGCCTGTG GAATTGGGGC TTGCTATGCC TGC GTTTGCC	1140
ATAAAAAGG AGACACTACT GGAGCAAAAA GTGTCAAGGT CTGTGATGAA GGTCCGATTT	1200
TTAAAGCAAG TGAGGTTATC TTATGATGAA AAAyCCTTTA GCCGTTTCAA TTCCAGGTTT	1260
AACATTAAAA AATCCAATTA TTCCCGCCAG TGGTGCTTT GGGTTCGGGG A	1311

## (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6948 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TCCTCTGATG AACCGATTGA AAGAATTAAG GAAAGAA <sub>n</sub> AA AATATTACTT TAGTTGAGTT	60
AAGTGAGGAG TTAGGTATTC CACGTTCCAC TCTTAATAGG TATGAAAACG AAGATAGCGA	120
ACCAAAACAA GAAaCTTGGG AAAAATTAGC TGATTATTAT GGTGTTTCTA CGGCTTATTT	180
AATGGGGATA TCCAACCAAA AGGTTAGCGA AGAAAAAGCT TTGACGGCCG CAAAGAAAGT	240
TTATCAAGTC TATCTTTCCG ACGACGATTT AGGAAAAGAA ATTCGAAAAG CTCTAATGTA	300
TTTTAATAAA AATGATTTAG ATAGTGTCTT AAAACAAGCA ATGCAGCAGT ATTTTACTAT	360
CCCAGCCGTT GAATGGAATA CAGAATTTCA ATCTTTAAAA AATACTGGTT TTCTATACAG	420
TTGGTTAGAA GGATATTTGG TAGACAGGTA TCGAAAAGAA GTGAAACTA ATCATAATTT	480
AATTACAAAT ACGTATTATA GTATCCCTTC TATTGATGAT GTGAATGAGT ATGGAATTTT	540
CTCTGATGAA ATTCAAATAC CGTTGACTGA AGATTTTAA <sub>n</sub> CAAATTTTCA AAGTTGAGGA	600
TCTATCCACG AAAACACATG AAGAAATAAA AAAACTTTTT AAATCCGATT CTACGCTCAT	660
TGcTTACAGC TATGAAAGCT CCGTAGACGA TAAATTAAAA GATGATATTA ATATGATTTT	720
AGATAATGCT CGTTCCGAAA TTCTTAATCT AAAAGAAAAA TATCCTGATA AACCTAGTAG	780
AATAGaACAA ACAACCGTAC TCATATCTAA AGACGAAGAT ACTTCCTTTT GGTCAAGGAA	840
TGGCTCAAAT GATTATATAG ATGAATTGAA CTTATCTGAG GAAACTAAGA AATTATTTGT	900
TCATTTAGGT TCAGTATTTA TTGAGGAAAA AAAACGTAA <sub>n</sub> TCAAAAGCTA AAGAATGAGT	960
TTTTTTCTTT GTATAATCCA GTTATAGAAA CATCTCTCCC CTAACATCTA CATACTATTG	1020
GCGTGGTGGT AGATAGGAGA AAAACAATGA CACAAATTAA CAAGTATACA AAAAAAGACG	1080
GTACAACCGC CTATATGTTT AATGCTTATG TAGGACGTCA TCCCAAAACT GGTAAAAACG	1140
TTTATCGTAA GCGGCAAGGA TTTAAAACAA AAAACAAGC ACAAATTGCA CTTGCTGAAA	1200
TTCTGAAGGA TATTGAAGAA AATGGACTAG ATAATAAACC CTCTGTTTTG ACGTTTAAAC	1260
AGCTTTACGA GAAATGGTTA GCACAGCAAC GGTTAACGAT TAAACCATCT TCAATCGCTG	1320
TGAATAAGCG TTTTGTGAA AAACACATAT TACCTTATCT AGGTGACTGC AAAGTTGACG	1380
AAATAACCGT TATACAGTGT CAGGACTTAG TTAATAAATG GTTTAATCAA GGACACAAGC	1440
AATACTCGTT CTATAGAAAA CTCCTGCTC AGATTATGCG ATATGGCGAA TCTATGGAAC	1500
TAATGAATAC TAACCCAATG AGAAAAACAA TCCTTCCTAA GTGGAAAGAG GAAGAAACAA	1560
AACTGGAGTA TTACACTAAA CAAGAATTGA ATCATTTTTT TGATTGTCTA AAACAATATG	1620
GTAATTTCAA ACAGTTAGCA TTTTTTAGGT TGCTAGCCTT TACAGGTTGT CGTAAGAGTG	1680
AAGTTCTAGC CTTGCAGTGG AAAGACCTAG ACATAACAA <sub>n</sub> CAAGAGCGTA TATATAGGCA	1740

AGACCGTAGC	ACAAGACGAA	TTCTATAATA	TACTCACTCA	AACGCCCCAA	ACAGCTTCTA	1800
GCACACGTTC	TATCAGCTTA	GACAATGAAA	CACTTCAAAT	TATGGCTAAA	TGGCGTACAA	1860
TCCAACGTAG	TGAcTATTTC	CAAATGGGCT	TTAATACATC	TAGTGAAGAA	CAATATATTT	1920
TCACGAACAA	CCATAATAAA	CTGCTCTCAC	CAAATGTAGT	TAATATATGG	CTCAATTGCT	1980
TAATAAAAAA	ATATGATTTA	CCTAGTATTA	CACCACATCA	TTTtagGCAT	ACACATGCAA	2040
GTTTACTTCT	TCAATCAGGA	GTGCCTATTA	AAGAAGTTGC	TGAAAGATTA	GGACATACTA	2100
GCACTGCTAT	TACTGATCGT	ATCTATTCTC	ATGTTATGCC	TGAAGAAAAA	GAAAAACAG	2160
CAGATAAGTT	CGCTCAATTC	GTCGGTTTTT	AATAAATGCA	CAGACAAAAG	CACAGATCTT	2220
TACTATTTTT	ATGTAAAAAA	AAGAGCTAGA	ATCCTTATAA	ATAAAGGTTT	CTAACTGGCA	2280
TTAGTAACGC	TATTAATGAA	ACAACGTGTT	TAATCATATA	ATAAACCAAC	AAATAAGTAA	2340
ATATCAAATT	CTAAATTCTG	TAAAAAACGA	ACAATCACGA	TTAAAAAAAG	TATTCCTTTC	2400
TAGGAATACT	TTTTAAATTA	AATCGTTTGT	GTGGTCGGGC	GAACAACCAT	TTCCTGATT	2460
GACATACGAT	CTGGCGTGTC	TATGGCAAAG	ACCACTGCTT	GCGCAATGTC	TTCCGCCTTT	2520
AAGCCCCAAC	TTAACTGTTC	TAGATGTAAC	GTTTCAGCTA	CTACTCGATT	AGAAATTGTT	2580
TGATAAAGTT	CTGTTTGTAC	AGCACCTGGT	GAAATAATCG	TTGATTTAAT	GTTATTCTCT	2640
CTTTGTTCTT	GTCGCAATCC	TTCCATAATT	GCTCGAACAG	CAAACTTTGT	TCCACAGTAG	2700
ACAGCCGAAT	CTGGATAAAC	GACATGTCCT	GCCACAGAAT	CAGTAGCAAT	AATCTGCCCT	2760
GATTTTTGTT	CAACCATAAT	TGGAAGAACT	GCCGCAATGC	CATTTAAAC	ACCCATAATA	2820
TTAATATCTA	GCATTTGACG	CCATTCTCCT	TTTGGTGCTT	CAATAAGAGG	CGCCGTGGC	2880
ATAACTCCTG	CATTGTAA	AAGAACATCA	ATTCTGTCAT	ATTTTCCAT	TGTAAGCTTG	2940
ATTACACGCT	GAACCTCCTC	TTCTTTCGTT	ACATCTGCTT	GTTGCACGAG	AATCGTTGCT	3000
TCAGGAAGTT	CTTTTTTTAT	AGCAATTAAA	CGTCTTGTC	GACGTGCTGC	AATGACTAAC	3060
TTGCTCCTT	TTCTGGCAAG	TAAACGGGCT	GTTGCCTCAC	CGATGCCACT	GGAAGCGCCC	3120
ATGATAACAA	TAACTTTTTC	TGATAATGAT	TTCATTATAT	TTCTCCTTA	TCTTCCTGAA	3180
TCCAACGAAT	AATTTTAGCT	GGTGTTCCTG	CAACAATAGC	GTTAgcTGGc	ACATCTTTAG	3240
TTACAGTCGC	ATTAGCTGCT	ACGATAGCAT	TTTACCAAT	CGTAATACCT	GGTAAAACAG	3300
TGACACCCGC	GCCTAACCAC	GCAAATTTTT	TAATGGTAAT	TGGTGCTAGC	AATACGCCTC	3360
GCCGTTTCAT	TGGCAGTTCC	GGATGATTGA	CACCTAAGAG	ATTAACGCGC	GGACCAATTA	3420
AAACATTGTC	TTCAATTATA	ATGCCCCCTA	AATCAACAAA	CATAGCTGCC	CGATTAATAA	3480
AAATATTTTT	ACCAAATGT	ATATTGCGTC	CAAAATCTGT	ATAAAAAGGT	AATAAATCG	3540
TAACGGTTTC	GTCAATTTTG	TCTTGATAA	TCGTGGCAAG	AAGATTCCTT	ACTTCATTAT	3600
GTGTATAGGC	TTGTTGATTT	AATTCTATTA	AATAGCGACT	ATTTTCTCT	TGAATCTGAT	3660
GAATAAGTTG	AAATAGTGGT	TCATCTTTAT	CAATGCTACC	TGTTGCAATT	AGTTCTTGGA	3720

GATTTTTTGT AAGCATTCCCT AATCGCCTCC TTTTGCTAAA TTTATCGTAA TACGTGACAA	3780
CTCCTCTGTC TAATACTTAT AATAAATAGA TAAGTATGCG TTTTAGTTAT AGGAGGGAAA	3840
GTAATGGAAT TACGAGTGAT TCACTATTTT TTAGCAGTGG TCCAAGAAAA AACGATTAGT	3900
GGCGCAgcCA AACCAATTGCA TGTATCACAA CCAACATTAT CTAAGCAATT AAAAGAATTA	3960
GAGGAAGAAT TAGGCGTGAC ATTATTTATA CGAGGAAATC GACAAATACA ATTAACACCT	4020
GAAGGAGAAT ATTTAGCTAA ACAAGGGCAA GATATTTTAA GCTTAGCGAA TAAAACAGTC	4080
ACCAACTTGT CGCAAAATGA ATTCATTAAT GGCGAAATCA CGATTGGTGG CGGCGAAACA	4140
AAAGCTATGT CTTTTTTAGC GAATGCACTA CAACAAATAA CAAGCCAGCA CTCAGCTGAT	4200
ATTCACCTTC ATTTATACAG TGGGAATGCA GATGATGTAA TAGAACGGCT AGACAAAGGA	4260
TTATTAGATT TTGGTTTAAT TATTGAGCCT GCACCTAAAC AAAAATATAG CTATTTAACA	4320
TTACCAGTTG TAGATACATG GGGCTTAATT ACTGTAAAGG ACCATCCCTT AGCCACTAAA	4380
AATGTTATTA CTGCAGCCGA TTTAAAAGAA GAACCTTTAT TTATTTCTCG ACAAGCACAA	4440
GTTCCGAGCC AACTCTCTGA TTGGCTCGAA GCAAGTCTAG ATCAGTCCG AATCGTTGGG	4500
ACCTACAAC TACTTTACAA CGCTTCACTG ATGGTAGAAG CTGGGCTAGG TAGCGCCCTA	4560
AGCATTGATG GTATTCTAGA AACAAAGCAA ACAAATTAC GTTTTATCCC GCTATATCCT	4620
GCTTTAACAG CCAAATTAG TTTAATTGG CGCAAAATA CGGTTCTTTC CACAGCTGCT	4680
GCATTATTTT TAGAACAAAT AAAAAAAGT ATTCAACGCC CTGAATAATA GGTCGTTGAA	4740
TACTTTTTTA AATCATCACT GTGATCGTAG CTGCTGCCAA AATCAAACT AAGCCGATAA	4800
TCGTTACGGT CATTCCTTG GCTGTTTCT TTTGACCTAA GAACCAATA CCTGTCAACG	4860
TTGCAAGCAC AACTGATGTT TGGGATAGAA TAAAGCCAGT TGCCAAGCCG TTCATATCTG	4920
GcTGCGCTGA AATTAAATAA GTTAACGCTG CAAAGGCGAA GAAAAGCCT GAAAAATTT	4980
GTTTGTAAAG AACGGCTTCT ATAAATGGGG ATATTTCCCC ACCCTTAATA GTGACAACCG	5040
CTGAGTAAAT AACTGCGACA ATTACCATGC CTATTGCTTG TGGTAAAAG GCATGCATGC	5100
CATCAATTGC CGTTGCTTGA GGTGCTGCAG AGTACGCCCA ATAGCCGATT TCACCAACGG	5160
CTAATAGAAG AACTGCTTTT TTCATAATCC CAGCACTCTC AGCGGACTCC GTTCCGACC	5220
AGACCGTCAT TTTGGCACCA ATCATGATTA AACTAAGGC AAATGCGCCT AACCAATTGG	5280
CTGTCGCACC TGGCCAGTTT CCTAAGAAAA AGACACCCCA GAGAGAAGCA CCTAAAAGTT	5340
GGAACGCGGT CGTTACTGGC ATGGCACGTG ATGAACCAAT CATCGTGAAG CATTTAAACG	5400
TAATGATTTG CGCACAGGCC CAACCGACAC CAGATAAAAT TGAAAAGAAT AAATCCATGC	5460
CCGCTGGAAA AGCTAAACCA TTAATCATTG AAAAAATAGC GGCAAAATC AATGTTCTTA	5520
AACTGGTACC GAGAATTTGA TTTACTGGAC GACCACCAAT TTTGGAAGCG ATTGTTGGGA	5580
AAAGCCCCCA ACCTAGAAGC GGTCTTAAAC CGATTAATAA TGCAGTCGCA TTCATCTCTA	5640
TTTCCCCTCT CTCGTTTAAA AGACAACGTT ACTTTCTAAT AAAATATTTG CGTAAGGAGT	5700



CATTTCCCCT	GTCCGAATAA	ATGCATGGCA	GTTGTTTAAT	TCTTGTTTCA	TTTCTGTATG	5760
AGGAATAAAA	CTGATGGGAG	TTTCTGGTAA	ACGAGTTTCA	ATTGCTGCCA	ACATGTCAGG	5820
ATTTTCTGTT	TTAATTTCTT	CTGCTAAATA	AATTCGTTGA	ACTGCCAATT	CCTCAAGGAC	5880
ATTGTTTAAA	ACTTCCATAA	AACTAGGGAC	GCCATTGGTA	ACTGCTAAAT	CAATTTTTTC	5940
TGTAGTCCGA	GGCACTGGCA	TGCCTGCATC	ACCAATACTT	AATTTATCAA	AATGACCCAT	6000
TTGTGCGATG	ACGCGGGAAA	TATCTGAATT	AATAACTTTT	GTTTTTTTCA	TGTCGCTCGC	6060
TCCTTTTTAT	TTGTCTGCCA	ATTCGTGTTG	ATAAGGAATC	GAAGGTTGGG	CTCCAAAACG	6120
TTGAACAGTC	AACGAAGACG	CCTTGTTTCC	ATAACGAATA	GCTTCTTCCA	AATTGCTAAA	6180
ATCTTTTTCT	AATATACTAC	TTAATGCGCC	AATAAAAGTA	TCGCCAGCAG	CCGTTGTATC	6240
AACCGCTTTC	ACTTTAAAAG	CAGGCACAAT	ACCACTTCGT	CCATTGACGT	CATAAAAGGC	6300
GCCTTTACTA	CCTACTGTAA	TAATTACTGC	TTCAAYCCCT	AATTGATGAA	GTGCTTCTGC	6360
GGCTTTACGC	ATACTCGCTT	CaTCTGTGAT	TTTAATGCCT	GTTAAATTT	CGGTTTCTGT	6420
TTCATTTGGT	ACAATCATAT	CTGTCACGTT	TAGTAATTCT	TCAGGAAC TT	GTTCTAACGC	6480
AGGTGCGGGA	TTTAAATCG	TTTTGACGCC	TGCTTTTTTA	GCAATTTTAA	ACGCTGCAAT	6540
CGTACTATCA	ATCGCACTTT	CAAATTGGGC	AATCACAAAA	TCACTTTTTT	CAATAATTTT	6600
TTGATGTTCC	TGGACTTGTT	TTGGCGTAAA	GGCGTTATTG	GCTCCTGCGT	AAATCATAAT	6660
GCTATTTTCA	CCAGCGTTAT	CTACCAyAAT	AAAGGCTTGG	CCAGTAGCTG	TTTCTCTAA	6720
AGTGGTGACC	CCAGTTAAAT	TAATTTCATC	TTGACTCATT	AAATCAGTCA	TCATAGCTCC	6780
AGCGCCATCA	TTTCCACAG	CACCAATAAA	ATATGTTTCT	GCGCCTGAAC	GTTTCGCTGC	6840
AACTGCTTGG	TTTGCTCCTT	TACCACCACC	TGCTGTAAAA	TGTTCAATGG	CGTGAATTGT	6900
TTCGCCTGGT	TTTGGCATT	CTTTGACTCT	TAATGTTGTG	TCTAAATT		6948

## (2) INFORMATION FOR SEQ ID NO: 5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8486 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AATTTGTTGA	ACCAGCGATT	GCGCnGCnAA	TACTGGTTTT	TCCAGTGCCA	GGCGGCCCAT	60
AAAGAATCAT	CGATGAAAGC	ATTTTGGCTT	CGACCATCCG	ACGAATAATT	TTACCAGGGC	120
CGACTAAATG	TTGTTGACCA	ACTACTTCGT	CTAAGTGACG	AGGTCGCATT	CGATAGGCTA	180
AAGGTTGTTG	CATAGAAGGA	CCTCCTTTTT	CATAATCATC	TTATTATAAC	ATAAAGAACG	240
TACATyCGCT	TTTTGGTGGk	TTTTTGGGkT	TTAGATGGAT	ATAGCGGAAG	GTTcGTGtAT	300
GATAGGGAGT	GGTGAAAGAA	ATGAAAGAAA	CGATAAATAT	CTTTAACACA	AAAACAAC TG	360

AAGAAGTGGC ACAATACTTG TTAGGGATGT ATTTAGmGCA CGAAACAGCA ACAGGCGTTT	420
TAGGTGGGTA CATTGTGGAT GCAGAAGCTT ATTTAGGTCC TGATGATGAG GCGGCCCAT	480
GCTTTGGCTT GAGAAAAACG CCACGCTTAC AAGCCATGTA TGACAAACCA GGTACGATTT	540
ATTTATATAC TATGCACACG CATTTAATTT TAAATATGGT GACACAAGAA CAAGGCAAAC	600
CACAAGGGGT GATGATCCGC GCGATTGAAC CTGTTGAAGG TGTCGACAAA ATGATTGAAA	660
ATCGCCAAGG ACGTCAAGGC GTGGAATTGA CCAATGGACC AGGTAAATTA GTTGCAGCCT	720
TAGGAATTGA TAAGCAATTA TATGGGCAGT CGATTTTTTC TAGTTCGTTA CGGCTCGTGC	780
CAGAAAAACG AAAATTTCCG AAAAAAATTG AGGCACTACC ACGAATTGGG ATTCCCAATA	840
AAGGTCGTTG GACAGAGTTG CCTTTAAGAT ACGTCGTGGC TGGCAATCCA TATATTTCTA	900
AACAAAAAAG AACAGCAGTA GACCAATAG ATTTTGGCTG GAAGGATGAA GAAAATgAAA	960
AAAGCAACAA TGCTCACATA CTTAGAGGAA CAACTTGAAA AACACTTAGG AGATTACGAA	1020
gTAGGCCTTG ATTGGGATCG CAAAAACCAT ACCATCGAAG TCATTGTTcG TTTATATGCA	1080
GAAAaCAATG AGCmAGTGGC GATTGATGAT GTTGATGGTA CGCTTTCAGA AGwAGAATTC	1140
ATTGAATTTG AAGATGGTTT GTTATTTTAC AATCCGCAAA AGTCTGTCGT TGATGACGAA	1200
GAGTATTTAG TCACAATTCC TTATGAAGGG AAAAAAGGgC TACGCAAAGC AGTTTTAGAC	1260
GGATTCATTC ACTATTTAAA AGTGGTTTTA GATGAAGGGC AAAGTGATTT GTTAGACTTT	1320
TTATCAGATG AAACAGCGGA AGTTTTTGAA TTGCATTGGG AGCCAGCAGA TTTTGAAGCG	1380
ATGATTAAAA AAGTGGCAGA AACAGAAAAA GAACAATGGA TTGCGTATCC AAGTTATTAA	1440
TTTAGAGGTG AGAACATGAA GTGGACAGAA GTAAAAGTCG AAACGGCTAG CGAAgCAGTT	1500
GAAGCAATTT CAAACATTAT GATGGnArCT GGCgCAAGTG GCGTGGCCAT TGAAGATGCG	1560
TTGGACATTG AAAATTTTGA AAGTGATCTG TATGGGGAAA TTTTAGATAA AGAGCAATTC	1620
ACCCACATTA AAGAAGGGGC GATTGTGATG GCTTATTTTC CTGAAACAAC CTTCTTACCA	1680
GAAATTTTAC CATTTATGAA AGAAAATATT TTACGCTTGC CTGAGTACGG CTTATCCATT	1740
GGTAAAAACG AAATGACAAT TAGTGAAGTA GCAGAAAGCG ACTGGGCAAC TGCTTGAAA	1800
AAATATTATC ATCCTGTCCG TGTCACACGA TTCTTAACGA TTGTGCCAAG TTGGGAAGCC	1860
TATCACGCAC AAGATGAAGC AGAAAAAATT ATCACATTGG ATCCGGGCAT GGCTTTCGGT	1920
ACAGGCACGC ATCCAACGAC TCGTTTAACC TTGCAAGCTT TAGAAaCTGT TTTACGTGGT	1980
GGCGAAACAG TTCTAGATGT AGGAACGGGT TCTGGTGTtT tGAGTATTGC AAGTCGCTAT	2040
TTAGGGGCTA AAGACGTCTA CGCATATGAT TTAGATGAAG TAGCAGTTGC AGCGGCAAAA	2100
GAAAaTATGG ATTTAAATCC CATTGCGGCC GATGTTCATG TGTCAGCCAA TGATTTACTA	2160
AAAGGGATTG ACCATTCTGC TGATGTAATC GTCGCAAATA TTTTGGCAGA TATTATTGTG	2220
TTAATGATTG AAGATGCTTG GCGTTTGCTA AAACAAGACG GTACCyTCmt TAAtTCTGGA	2280
ATTATTGAAG wTAmAAAAGC aATGGTTTTA GAAGCACTAA CGAAAGTCGG TTTTGTGGTG	2340

GACCAACTCT	TTAATCAAGG	CGATTGGTAT	GCGATTaTCT	TGAAGAAACC	aGAGGAAGAG	2400
TAAaTGCA	CGGTATTTCA	TGAAAGAAaGA	TTATCCAgAA	AAgATTTGTA	TAAgTCGCAG	2460
ATGAAAATA	TCATCATATT	GTGCGAGTTA	TGCGAATGAC	GCCGAATGAT	CGTTGTTATT	2520
TAGTATTTCA	AAATAAAACT	GCCATTTTAG	CTGAGATTGT	GGAAATTGAT	TCAACATCTG	2580
TTTACTTTAA	AGAAATTAGT	AAAGAAGAAA	TGGACAAAGA	ATTGCCAATT	GAAAGTACGA	2640
TTGCGTGTGG	GTTACCAAAA	GGGGACAAGT	TAGAATGGAT	TGTTCAAAAA	GGCACGGAGC	2700
TAGGTGGCAA	TCAGTTTATC	GGCTTTCCTG	CGAAAAtTCC	GTTGTCAAAT	GGGACCACAA	2760
AAAAAGAGCA	GCCAAAGAAA	AAAGATTACA	AAAAATTGCC	ACAGAAGCGG	CGGAGCAATC	2820
GCATCGACAG	CAGACACCAA	GTGTTTCTTT	GGTGGAAAAA	ACACAAGAAA	TCATTGCGCA	2880
ATTTGACTCA	TATGACACAG	TCTTGGTTGC	GTATGAAGAA	TCAGCAAAAC	AAGGAGAAAA	2940
AAGTCAATTA	GCACAAGTAT	TATCCACTTG	CCAGCCAGGC	GCACGTTTGT	GTGTGCTCTT	3000
TGGACCTGAA	GGTGGCTTTG	CACCGCAAGA	AATTGAACAA	TTTCTGCAGG	CTGGGGCGAA	3060
ACTTTGCGGG	TTGGGTCCGA	GAATTTTACG	AGCAGAAACA	GCACCGTTGT	ATCTTTTAAG	3120
TGTGGTTAGC	TACCAATGG	AACTTTTAAA	TTAAACGAAA	GCCCGCTGT	AAAATTCTTG	3180
TAGGCATTCT	CTTTAAATA	ACAGAGAGTA	AAAAAGCGAC	AGTGATTGAA	ATTTAAGGGA	3240
AATTTAGTAT	AATAGAGTTG	TTGAAAGTGC	TACTGTTTCA	AGTGGTACTT	TTCTTTTTTA	3300
TTAAAATAGA	TAAACGTGA	ACGAATgnTC	TTTATCCAAA	GGTCACCTTC	ATAGAGGAGT	3360
GAATACCATG	CCAAAAGAAG	AAATACTGAC	AGGCCCAGCG	GTCATCAAAA	TTGTAAGTAC	3420
CTATATGGGA	CCAGAACATG	TTGAGTTGGT	TCAAAAAGCA	TTAACTTATG	CTGAAAAAGC	3480
ACACGAAGGT	CAAGTCCGAC	AATCAGGTGA	ACCTTACATC	ATCCATCCGA	TTCAAGTGGC	3540
AGGCATTCTG	GCTGAATTAC	ATATGGATCC	CCATACTGTA	GCAACAGGAT	TTTACATGA	3600
TGTTGTCGAA	GATACGGACG	TAACACTGGA	CGATTTAAAA	GAAGAATTTG	GCGAAGACAT	3660
CGCAATGTTA	GTCGATGGCG	TGACGAAACT	AGGAAAAATT	AAATATAAAT	CTCATGAAGA	3720
ACAATTAGCT	GAAATCATC	GGAAATGTT	ATTGGCAATG	GCGCAAGATT	TACGGGTCAT	3780
TATGGTGAAA	CTAGCCGATC	GTTTGCACAA	TATGCGCACA	TTAAAACATT	TACGTGAAGA	3840
TAAGCAACGA	CGGATTGCAC	AAGAACTTT	AGAAATTTAT	GCGCCTTTAG	CACATCGCTT	3900
AGGGATTAGT	CGAATTAAAT	GGGAATTAGA	AGATACTGCC	TTACGTTATA	TTAACCCCAA	3960
CCAATACTAT	CGAATTGTTA	ATTTGATGCA	AAGCAAACGC	GATGAACGTG	AAGCTTATGT	4020
GGCAGAAGCA	GTCGAAGATA	TTCGTTTAGC	yACGGAAGAT	TTAGAAATTT	ACGCrGAAAT	4080
ATATGGTCGA	CCAAAACATA	TTTATTCAAT	CTATCGTAAA	ATGAAGGACC	AAAAGAAACA	4140
GTTTAATGAA	ATCTATGACT	TATTAGCCAT	nCGTGTTATt	GTCGATTCCA	TTAAAsaTTG	4200
CTATGCCGTG	TTAGGGcGAK	cATACACGTT	GGACACCAAT	GCCAGGTCGC	TTTAAAGATT	4260
ATATTGCTAT	GCCAAAAGCG	AATATGTATC	AATCTATCCA	TACGACAGTT	ATTGGACCAA	4320

AAGGGAATCC	AGTAGAGGTC	CAAATTCGGA	CCCATGAGAT	GCACCAAATC	GCCGAATTTG	4380
GGGTGGCGGC	TCACTGGGCC	TATAAAGAAG	GCAAAGCGGA	AAAAATTGAA	ACAGATGAAC	4440
TGACGAAGCA	AGTCGATTGG	TTCCATGAAA	TTATTGAATT	ACAAGATGAA	AGCTATGATG	4500
CTTCTGAATT	TATGCAAGGT	GTGAAAGAAG	ACATTTTTAG	CGACAAAGTC	TATGTCCTTA	4560
CGCCAAGTGG	AGATGTAACC	GAAGTCCAA	AAGGCTCCGG	ACCACTAGAT	TTTGCTTACA	4620
GTATCCATAC	TGAAATTGGG	AATAAAACAA	CCGGTGCGAA	GGTCAACGGC	AAAATGGTAC	4680
AGTTAGACTA	TGTCCTAAAA	AATGGGGATA	TTATTGAAGT	CCTAACATCA	CCGAATTCTT	4740
TCGGACCTAG	TCGCGACTGG	CTGAAAATGG	TTAAAACAAG	TAAAGCCAAA	AATAAAATCA	4800
AACGTTTCTT	TAAAGAACAA	GACCGTGAAG	ATAACATTAT	TAAAGGGCAT	GATGCTGTCA	4860
TTAAATACAT	GACAGAAATC	GGCTTTACAC	CGAAAGAATT	TTTAACGAAG	AATAAAATGG	4920
CGGAAGTGTT	AGATAAATTT	AAATTTCAAA	CGGAAGATGA	TTTATTTGCG	GCAGTAGGTT	4980
ATGGCGAAGT	AAGTGCACAA	GTAAGTGTGA	ATCGTTTAAC	AGAAAAAGAA	CGCCGCGAAC	5040
AAGAGTTGGA	ACGCCAACGT	CAAGAAGCAG	AAGAGTTGTT	AAATCAGCCA	GCTAAAAAAG	5100
AATCAGAAAA	AATGAAAGTT	CGTCATGAAG	GCGGCATCGT	GATTCAAGGG	GTGGACAACC	5160
TATTGATTTC	TTTAAGCCGC	TGTTGTAATC	CAGTGCCTGG	TGATGAAATC	GTTGGTTACA	5220
TTACCAAAGG	ACGTGGCGTT	TCAATTCATC	GTGCGGACTG	TCCTAACGTA	CAACATCAAG	5280
AAGAACTAGC	ACAACGTCTA	ATTGAAGTTG	AGTGGGAAGA	TACCGAGCAC	AGTCGCAAAG	5340
AATACGCTGC	CGATTTGGAG	ATTTACGGCT	ATGATCGTAG	TGGCTTACTG	AGTGATGTTT	5400
TACAAGTCAT	TAGTTCGATG	ACAAAAAATT	TAGTCGGCGT	TGAAGCTCGT	CCATCCAAAG	5460
ATAAGATGGC	CCAGATTCAT	GTTACTGTGA	AAATTCAAAA	CCTTTCTCAT	TTAAAAACCA	5520
TCGTTGACAA	AATTAAGT	GTACCAGATG	TTTATAGTGT	TCGTCGAACC	AATGGGTAAG	5580
TGTTAAAAGT	GAGGCAAAAA	AGATGAAAGT	AGTTATTCAA	AGAGTCAGTC	AAGCACAAAGT	5640
TGCGATAGAA	GAGCAGATAG	TGGGACAAAT	CAAACAAGGC	TTCATGGTGC	TGGTGGGGAT	5700
TCATCAAGAA	GACACGCCAG	AAGATGTGGC	TTATGTTGTT	GGCAAAATCA	GTAAATTGCG	5760
TGTTTTTGAA	GATGACGAAG	GAAAAATGAA	TCGCAGTATT	CAAGAGATTG	AAGGATCAAT	5820
TTTGAGTATC	TCTCAATTTA	CCCTCTATGC	AAAACTAAA	AAAGGCAATC	GTCCCAGTTT	5880
TATTGAAGCG	GCTCGTCCAG	ATGTGGCGAT	TCCTTTGTAT	GAATTATTTA	ATCAACAATT	5940
AGAAGCAGAA	GGAATAGCGG	TAGCCACAGG	AGAATTTGGC	GCGGATATGC	AAGTGTCGTT	6000
GACCAATGAT	GGTCCAGTGA	CCATTGTGAT	TGATACACGA	GAAAAATAAT	TGATTTTCCT	6060
GTCCGAACCC	CTTGACTTTC	TAAGAGAACT	TTAGTAGTTT	ATAGATAATA	TTAAAAATCT	6120
ACGAAGGAAC	GAGTAAGTTT	CATCTGTCTG	CGAAGAGAGA	GTTGCCGTGG	CTGAGAGCAA	6180
CTTCAACAGT	GAAAGTGAAA	GACATTCTGG	AGATGGACGG	AAGAAAAGCC	GTCCCGTTAC	6240
GAGCGTTAAT	CGTTGAGGAA	AGAGTAAAT	CTTTCAAAAC	TAGGTGGTAC	CGTGTAAG	6300

AGACTTTTAC GCCCTTGTAT GTTCAACATA CAAGGGCTTT TTTCATTTTT AGAAAAAAGA	6360
AAATGGCGCA CACATTTTTG TGAATGATA GAAGGAGAGA AAGACAATGA GTTATCAAAA	6420
ACCAAAAGGA ACAAACGATA TTTTGCCAGG AACTTCTGAA AAATGGCAAT TTGTGGAAGA	6480
AACAGCTCGT TTGATTTTTA AAGATTATCA ATACCAAGAA ATCAGAACCC CGATTTTTGA	6540
ACATTATGAA GTAATATCTC GCAGTGTTGG CGATACCACA GATATTGTTT CAAAAGAAAT	6600
GTATGATTTT TATGATAAAG GAGACCGTCA CGTGACGCTA CGTCCTGAGG GGACAGCGCC	6660
aATTGTTCGG GCCTTCGTTG AAAATAAATT ATATGGTCCG GAATATACGA AACCATATAA	6720
AACCTATTAC ATGGGGCCGA TGTTCCGCTA TGAACGCCCA CAAGCTGGTC GTTTGCGTCA	6780
ATTCCATCAA ATTGGTGTGG AAGCATTTGG TAGTGAAAAC CCAGCATTGG ATGTTGAAAT	6840
CATGGCTATG GCTTTGGACT TCTTCAAACA ATTAGGCATC CAACAAATCA AATTAGTTAT	6900
TAATTCCTTG GGGGATAAAG AAACACGTGC TACGTACCGT CAAGCATTA TCGATTATTT	6960
AGAGCCCCAT ATGGCAGAAT TAAGCGAGGA TTCACAACGT CGCTTACACG AAAACCCATT	7020
GCGGGTGTTA GACAGCAAAG ATAAAAAGA CAAGGTGATT GTCGCAGAAG CGCCCTCCAT	7080
TTTGGATTAT TTAAATGAAC CATCTAAAGC ACATTTTGAA GCAGTAACTG ATATGTTAGA	7140
TTTACTAGAA ATTCCTTATG AAATTGATAG TAATATGGTT CGTGGCCTGG ATTATTATAC	7200
ACACACAATT TTTGAAATTA TGAGTGAAGC GCCTAAAATG GGTGCGCAAT CAACTATTTG	7260
TGCAGGAGGC CGATACAATG GTTTAGTTGA AGAATTAGGC GGCCAGACA CACCAGGTTT	7320
TGGTTTTGGT ATGGGCATTG AGCGAGTGTT GTTAACAATG GAAGCTGAAG AAGTTGTGAT	7380
TCCAGCGTTA TCTGAATTAG ACGCATATGT GGTGGGATT GGTTCAGACA CCAACGTCGC	7440
AGCTTTGCAA CTTGTTCAA GCATTGTA CTTTGTTTC TCAGCTGATC GTGATTACAT	7500
GAATCGCAA CCAAAGCGC AATTTAAAC GGCCGATAA TTACAAGCAA AATTAGTTTT	7560
AACAATCGGT GAAAATGAAT TGAATGAAGG CATTGTCAAC GTAAATCAA TGGCAACACG	7620
CGAAGAAAA GCCTTCCCGT TAAGTGCTAT TCATGATTCA TTTGATGAAG TGTATGACGA	7680
AATGATGACA AAAATGATTG AAGAATAAAA AGGAGAGATT GAGAAATGGA AAAACGCACA	7740
ACCTATTGCG GCAATGTATC TGCCGAATTT ATTGAAAAAG AAGTTGTATT AAAAGGATGG	7800
GTTCAAAAAC GTCGTGACTT AGGAGGCGTT ATCTTTATTG ATTTACGTGA CCGTGAAGGA	7860
ATCGTCCAAG TGGTCTTTAA CCCAGAAAAG TCAAAAGAAG CATGGGAAAT TGCTGATAAA	7920
TGTCGTAGTG AATACGTCAT CGAAGTTAAA GGTCAAGTCG TTTATCGTGA TAAAGAAGCG	7980
ATTAATCCCA AAATGAAAAC TGGCGAGTTT GAAGTGATGG CGACTGACAT TACGATTTTA	8040
AATACTGCGA AAACAACGCC ATTTACCATT GAAGATGACA ATAACGTTAA TGACGAATTG	8100
CGCATGAAAT ATCGTTATTT AGATTTACGT CGTCCATCAA TGACAAACAA CATTAAATTA	8160
CGTCACCAAG TAACGAAAAC AATCCGTCAC TATTTAGATA ATCATGACTT TTTAGACATT	8220
GAAACTCCTT ATTTAGGCAA ATCAACACCA GAAGGCGCTC GCGATTATTT GGTTCCTTCT	8280

294

CGTGTCCATG CCGGTCATTT TTATGCTTTA CCACAATCAC CACAACCTCTT TAAACAATTA	8340
TTAATGGGGG CAGGCTTCGA TCGCTATTAC CAAATCGTTC GTTGTTCCTG TGATGAAGAT	8400
TTACGAGGCG ACCGTCAGCC AGAATTTACA CAAATCGATA TTGAAACGAC CTTTTTAACA	8460
CCAGAAGAAA TTCAAACATA TACAGA	8486

## (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 789 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AATGTCAAAA TTATTAGTTG TTAAAGCACA TCCATTAACA AAAGAAGAAT CACGCTCAGT	60
TCGTGCGTTA GAAACATTTT TAGCTTCATA CCGTGAAACA AATCCATCAG ACGAAATCGA	120
AATTTTAGAT GTTTATGCCC CAGAAACAAA CATGCCGGA ATCGATGAAG AATTATTATC	180
TGCTTGGGGT GCACTTCGCG CTGGCGCagc ATTTGAAACA TTAAGCGAAA ACCAACAACA	240
AAAAGTGGCT CGTTTTAATG AATTAAGTGA TCAATTTTTA TCTGCAGACA AAGTAGTAAT	300
TGCTAATCCA ATGTGGAAGT TAAACGTACC GACACGCTTA AAAGCTTGGG TAGATACAAT	360
CAACGTTGCT GGAAAAACAT TCCAATATAC TGCAGAAGGA CCAAAACCTC TAACAAGTGG	420
TAAAAAAGCC TTACACATCC AATCAAATGG CGGCTTCTAC GAAGGAAAAG ATTTTGCTTC	480
TCAATACATT AAAGCGATTC TAAACTTTAT TGGCGTTGAT CAAGTTGACG GATTATTCAT	540
CGAAGGAATC GATCATTTCC CTGATCGCGC GGAAGAACTT TTAAATACTG CCATGACCAA	600
AGCAACTGAA TACGGTAAAA CATTCTAAAT TTTCCGAGTG ATTTACATCA AAAaGAGACA	660
AGCAGTTCGT TCACTGCTTG TCTCTTTTGT TACGTAGATT CTTGTTGaTC TGGcTTATCt	720
TCATcATACA AACCGGaAAC GaTTTGGtAC CACTCaAAAA TGTGTGTAAT ATCACnTTAG	780
CTGCGGCAT	789

## (2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4159 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CGTCGTTTAT TACGTCGTGA CGGTAATGCA CGGGAAAAAA ATTAGGCATT AATGAAGCCT	60
TTTTATACAA .ATTAGTGCCT GTTGTGGCG AAATCATGGT AAGCTACTAT CCAGAAGTAT	120
TACAACAAAA AGACTTCATC GAAAAAGTGG TTCGGACAGA GGAAGAAngT TTCCATGAAA	180
CCATCAATGA AGGTTTAAGC ATGTTAAATG AGGTTATTAA AGAAGTTAA GATGCAAAAG	240

GCGATACATT AGATGGAAAA ATTATCTTCA AACTTTATGA TACTTTTGGc TtCCCCGTAG	300
AATTAACGGA AGAAGTTGCA GAAGATGAAG GCTtGAAAGT CGACCAtGCT GGGTTTGAAA	360
CAGAAATGGA AGCACAGCgT GAACGTGCTC GTTCTGCCCC CAGTAAAGAA ACTTCAATGG	420
GTGTTCAATC TGCTTTATTA ACAGATATTA AAGTAGAAAG TAAATTTGTT GGCTATACAG	480
AATTAACACA TGATAGTGAA TTATTTGTTA TTATTCAAGG TGACGCACTA GTAAATGAAG	540
CATCTGCAGG AACAGCCGAA TTAATTTTTG CTGAAACACC ATTTTATGCT GAAATGGGTG	600
GACAAATTGC TGACCGCGGC TATGTAAAAA ATACCGCAGG GgAAGTCGTT GCCAACGTGG	660
TGGATGTGAA AAAAGCACCA AACGGTCAAT TTTTACACAA AGTAGAAGTT TTGGCGCCAT	720
TAGCAGAAGG TCAAATTTAT CAATTGCAAG TGGACGAACG GATGCGGACA CGTaTTTTGA	780
AAAATCATAC GGCAACCCAT TTATTGCATC GTGCTTTAAA AGATGTGTTA GGGGaGCATG	840
CCAACCAAGC AGGCTCATTa GTTGCACCAG GACATCTACG TTTTGACTTT ACTCATTTTG	900
GGCAAGTTAC ATCAGAAGAA CTAGCGCGGA TGGAAGCCAT CGTTAACGAA AAAATCTGGG	960
AAGCTATTCC TGTTGTCACA ATTGAAACAG ATATTGATAC AGCGAAAAAC ATGGGCGCAA	1020
TGGCGTTATT TGGCGAAAAA TATGGCAAAG AAGTCCGTGT AGTTAATATT GGGGATTACT	1080
CTATCGAATT ATGTGGTGA ACACACGTTG CTAATACAGA AGATATCGGG ATTTTCAAAA	1140
TTGTTTCTGA ATCTGGAATT GGTGCAGGAG TACGCCGTAT TGAAGCAGTG ACGAGTAAAG	1200
AAGCGTATCA ACTTTTACAA GAAGAAGAAC GTCAATTGAA AGAGATTGCT ACATTAGTTA	1260
AATCACCTCA ATTAAAGAA GTGGTTACGA AAACAGAGCA ATTGCAACAA CAATTACGTG	1320
ATCTACAAAA AGAAAACGAA CAATTGGCAG GCAAATTAGC GAACCAACAA GCTGGTGACA	1380
TTTTTAAAGA TGTCAAAGAC ATCAATGGTG TTCGTTATAT TGCCGCTCAA GTCAATGTTA	1440
AAGACATGAA TCAATTACGC CAATTGGCTG ACCAATGGAA AAAAAAGAA TTGTCTGATG	1500
TACTAGTTCT AGCAACTGCA CAAGATGAAA AAGTAAGCTT GTTAGCGGCT ATGACAAAAG	1560
ACATGAACGG AAAAGGCTTA AAAGCCGGTG ACTTAATCAA AGCAATTGCG CCAAAAGTTG	1620
GCGGTGGCGG CGGTGGTCGT CCTGATATGG CTCAAGCTGG TGGGAAAAAT CCAGCAGGGA	1680
TTGCGGATGC CTTAGCAGAA GTTGAAAAC TGGTGGCAAA TGCCTAATTA AAAATGTTTT	1740
CCCTATTATT AATATAAGGA GAATGCTCTT TATAGCaATA ATGACAGCCC TCTTCAAAGT	1800
TATGAAATTG ATTTTGAAGG GGGTTCTTTT TTTtAAATGA TTAAGGAGAG GAAACAGTAG	1860
ATTGCACGAA TGTGTAAGTT TTCATCTTCA ATTGACAATG GTATAATAAG CAACGGAGGA	1920
ATCAAAAATA TGAATGAAAA TCAATTaTca AAACGTCTCG CAACGGTTGG TGATTTAATC	1980
CCTAGAGGAA GTCGCTTAGC AGACATTGGT TCAGACCATG CTTATTTACC AGTTGCTTTA	2040
ATGTTAGAAA ATAAACTGTC TTTTGCTGTG GCTGGGGAAG TGGTGAAGG CCCCTATCAA	2100
TCTGCCAAAA CCCAAGTCAG TAAATCCAAT TTGACGGATA AAATTATCGT ACGTTTAGCG	2160
AACGGGCTAG ATGCAATTGA ACCTGAAGAT CAaATAGACG TCATTAGTAT TTGTGGAATG	2220

GGCGGCACAT TAATCCGAGA TATTTTAGAA GCTGGACGCA AAaAAAATCG TTAAACAGGG	2280
AAGGAACGCC TTGTTTTACA ACCGAaTATT GGTGAACCAA CTTTACGCCG CTGGTTAATG	2340
GCAAATGACT ACAGTATTAT TGATGAAACA ATCGTGGAAAG AAAACCGTAA ATTGTATGAA	2400
ATTATCGTAG CCGAAAAAAC AGAGCAATCC GTTTCCTTATA CGGACCAAGA ATTACTATTT	2460
GGGCCAGTTT TAATTAAAAA ACAAGGACCG GTTTTTACTA AAAAATGGCA ACGAGAATTA	2520
AAACAGCGCA AAAGTGTCT TGCACAATA GCAAAAGCAA GTGGAGAACA TATTGAAAAG	2580
CAAGCGAAAC TACAGCAAGA CCAACAATTA ATTGAGGAGG TACTGGCCAA TGGCTGTGAA	2640
CGGTAAGACA TTGATTCGAC GCTTCAATGA CTATTGTCCC GAATGGTTAG CAGAAACGGG	2700
AGATCCTGTG GGCTTGCACA TTGGTACATT AGATAAACCA ATTGAAAATG TGATGGTTAC	2760
CTTAGATGTT CGTCTGAGG TCGTTGCAGA AGCAATTGAG AAGCAGGTTG ATTTAATTAT	2820
TGCGAAACAC CCACCTATTT TCCGCCAGT TAAGCGATTA ACCACAGATA ATTTTCAAGA	2880
AAAAATGTAC GCTGATTTAT TAAAACACGA TATTGCTGTT TATGCAGCGC ATACTAACAT	2940
GGATATTATT GACAATGGGC TAAATGATTG GTTTTGTGAA TTATTGGGAA TTAAACAGAC	3000
CACCTTTTTA ACTAAACAC ATACTGTTCC TTATAAAAAA TTAGCCGTTT TTGTGCCAAT	3060
TGATGAAGCT CCGCAAATGA GAGAAGCTTT AGGTCTAGCT GGCGCCGGGT CTCAAGGTGA	3120
TTATTCTAAA ACAAGTTACT CTTAATTGG AACAGGACGT TTTACACCAA CACAAGGAGC	3180
AAATCCAACG ATTGGCGAAA TTGGTCAAGA AAGTGTGCTC CAAGAAGCAA AAATCGAAGT	3240
GATTTTCCCT GAAACTAAGC AAGAACAAGT TCTTGCAGCT ATGTTACAGG CGCATCCTTA	3300
TGAAGAACCG GcGTATGATG TTTATACCAT TGAAAATCAA TCAAAGAAT TTGGTTTAGG	3360
TCGGGTGGA GTATTAGACA AGCCTGTGAG ATTCTCTTAT TTTGTACaGC AAGTCAAGGa	3420
AGCTTTCCaA TTGGATGGCT TCGGgkTCAT TGCAAAAGAT GATACTAAAA TGATTCAACG	3480
GGTAGCCATT TGTGGTGGA GTGGTGAAAA ATTCTATCAT GATGCATTAC GTAAACAAGC	3540
CGATGTCTAT ATTACTGGTG ATGTGTATTA TCATACGGCA CATGACATGA TTGCTGAAGA	3600
TTTACCAGTT ATCGATCCTG GTCATTACAT TGAAGCGCTT TGTAAGCCAr AACTTGTGGA	3660
ATTAATGAAT CAATGGAAAC mAGAAAaTGA GTGGgCTGTA TCCATTTTkG AaTCAGAAGC	3720
GAATACCAAC CCATTTCTGT TTAAATAAAA AAATGAAGAT TAAAGGAGTG TTCTTTTATG	3780
TATGAAAAATC TATTACCTCG TTTTTTACGG TATGTGAAaA CAGAAACACG CTCGGATGCA	3840
ACTAGCACAA CGACACCATC AACACAAACA CAAGTAGCCT TTGCACAAAC CTTAAAAAA	3900
GAATTAGAAG AATTAGGGAT GAGCGATGTT ATTTATAACG AAACAAATGG TTTTGTGATn	3960
GCTACGTTAC CTAGTAACGT GGAGAAAGAC GTTCGTTCAA TCGGCTTTAT TGCCCATATG	4020
GATACCGCTG ATTTTAATGC AGTGAATGTT TCTCCGCAA TTGTTGAGAA CTATGACGGA	4080
GAATCAACGA TTCCTTTAGA TAAAGAAGGC AAATTCACCT TAAACACGAA AGACTTTCCT	4140
AACTTAAAAA ATTATCGTG	4159



(2) INFORMATION FOR SEQ ID NO: 8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8377 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

ATCGATACAA TTTTCTTTT GTCACAACAG AAGAAGAATT ATTAGAGATT TTAAAAGATG	60
CCTTAAAAGA TAAAGCCTTA GTTGTCTAGTA CATTAGTCAG TAAACAACATA ATCACAGCGG	120
CGAAAGAATT TAGTGAACTGA ACAGGGTTGT TATATTTAGA TTTAATGGCG CCATTTTTTTG	180
AATTAATTCA AGCGAAAGCC GGAGTAGATC CTATTGAAGA GCCTGGACGA CGCCACCAAC	240
TAGATCGTGC CTATTTTGAT AAAATCTCAG CGATTGAATT TGCTGTAAAA TATGATGATG	300
GCAAAAATCC TCAAGGGTTT CTTGATTCTG ATATATTGTT GTTAGGCGTT TCGCGGACCT	360
CAAAGACGCC AGTCAGTATG TATTTAGCGA ATCAAGGCTA CCGCGTTTCT AACTTACCAT	420
TAATTCCAGA AGTTCCATTG CCGCCAATTT TGGAAGAAAT GGATCCACAA AAAATGATTG	480
GTTTAGTTTG TTCGCCAGAA ACATTAGGAC AGATTCTGAG CAGTCGGTTG GCTTCCTTAG	540
GTTTAGGTAA TGAGACCACT TATACCAATG TTGAACGGAT TGAACAAGAA TTAGCTTATG	600
CCGAAGAGAT TTTTGCGAAG TATGGCATCC CAGTGATTGA TGTAACCGCA AAATCTGTCTG	660
AAGAAACAGC CTTTTTAATT AAAGAAAAAC TAGATGAAAG AAATTAAGAA GCTTGGCAAC	720
CTCACTAAAT TTCAGTGAAG TTGCCAAGCT TTTTATACTT AATTGAGCTT CTTTTTGTTA	780
GTCAATATTT TTTTGATCGA CAATGAGTAA AGCGATCATC ACGTAGACAA TCCAACACT	840
TTTAGAATAA AGTGTATAGC GGGGTTGCCA TTTGCTAGCA AATTTTTCCT TGTATTCGCG	900
TAATCCTTGG AAGGAGTAAA TTTCTGAACC AAATTCATAA ATTAACGCAG CAATCCGTTT	960
TTGAATAAAA CTTTTTCTTG ATTGACCAAC GTTGGCTAAC GGAGCCATCC CTAAATTGAA	1020
ATAATGAATA TTTTCTGTTT TCATATATTC AAATAAATGA ATAAATAAGA AATCCATACT	1080
TCCAGAAGGG GCTTTTTCTT TATGGTGACG CATCAAGTCG ATGGTACCAA CTTCAATTGT	1140
GTAAC TAGGA ATGATATTAG CGAAAGAGAC CATTTCTCCT TCAGCATTTT GGACAACAGC	1200
AATTGGGTTT CTTTGAAGAT ACGCTTCATC GAAAAATCCT AGTGAAAAGC CTTTTTCTTT	1260
TCGATTGTCT AACCAACTAT CCGAAATAGC TTTTAATTCTG TGCATTTGTT CGGTACTAAA	1320
AGGAGGTTGG AGTACGTCAA ATGAATAGCC GGCTTTAGTA ATTTTATTAA AACTGCGCG	1380
GGTACCTTTG AATTTTTTCC CAGACATCGT AAATGTTTCT AGATTAACTA AGGCTTCTTC	1440
GGCCATTTTG ATAAATTCGT AGCCATATTC GTGTAGTAAC ATCACGCTGT CCTCATTGGT	1500
TTCGTAAAAG ACGGGTACAT AGCCGAAACG ATCTGTTTCT TTAAGGAATG CATCAATCGC	1560
CGCAGGAAAA TCTTCTTTGT TCCAGAAGG GTTGCCCATC ACAACGCATT TATTATTAAT	1620
GGTGTGAAT TGTAAGAAAG CAGTCGGTTC TTCTTTGGGA TATAAGAAAA CCTGTTTATC	1680

ATGAAGAAAG	ACAAGTTCAC	TGTCAATATT	CCCGCCATAA	GTTGTTAAGA	TATTATGCAA	1740
GATACTATCT	TCTGGAAACT	CACCAACTTG	GTGTTTTTTC	CCTTGGAGGT	AGCGAACAAA	1800
TAGAAAGTTA	AAACTCATAA	CTAAGAGGAT	AGCGATAAAT	CCTGACAACC	AGATCCGTTT	1860
GGAAGGGAAC	AAGAAAAATT	CCACAAAATG	ATGACGATGG	TGTGGGAAGT	TCGGTAAATT	1920
ATAACACCA	ATTACCAGGT	ATAACAAAGT	CAACAAGCCA	AAGATAAAGC	CATCAATCGT	1980
CATCCACTCC	CAAGAGTAAA	CCAGTTGTTC	ACGGTACAAT	TCTGATTTTG	AAAAAATAAC	2040
GATAAATAAT	AAAATAGCTA	AATAGAAAAT	GGAAAACATA	CTAAAGTCAA	CAACAAATGA	2100
ATACAAGATA	GCACCAATTA	ATAAAATAAT	CGTTGGATAA	TAGGCTCTTT	tAaCmCGATC	2160
AGCATCCCcG	TCCCATAATT	mATAATGCAA	AGCCTAAAAC	AATACTCGGA	ATTTGnATAA	2220
TAATATGTGA	ACGGAAGGGG	TTAATATCTC	TTAACCAATG	CACTTGAGTG	AAGGCTTCTG	2280
GAATAGTGGC	TAGGAGCACC	ATCATAATTC	CTGAAAAATA	AAGTAAAACC	ACTTCTAGCT	2340
TGTGGGCAAT	TTCCAAAGTA	ATTTGTTTTA	GCAGTTGACG	ATAATGAGTA	TCCAACTTTT	2400
GGCTTAAGTG	TTTTGTAAAG	AACAAACAGC	CAATTAAGAA	CGGAATAATA	TAATAGAATA	2460
AACGATAGAG	CAATAACCAC	AAAACAATAA	TTTCACGATC	GACACCTAAG	TTGCTCAGTC	2520
CTAAAATCAT	CATGACATCA	AAACTACCTA	AAGCCCCTGG	AATCATCGAA	GCAATCCCAA	2580
TAATCGAAGC	TGCTACATAT	AAAGGAACAA	TATCAATTAA	AGGAATCGGT	ACATCTAATA	2640
AGTAACCAAT	ACTGATAAAG	GTAATTAAAA	CACCTGTCCA	TTCTAAAAAC	GaGAAACAA	2700
CTAGAGATAA	GCGATTTTTT	GCATCGAGGT	TACCAAAGGA	ACTATTTTTT	TGGAAATGTA	2760
AAATAAGATA	AAGTAACGGG	AAATAAAGAC	CGCCACCTAA	TAACCAAATC	CAATATTGCT	2820
GTAAATAAGC	CATCTTTGGA	TCAAAC TGAA	CAAGGAAAAA	GCAAAGAATA	CTATAAATAG	2880
ATAGACCGGA	TAACACAAAA	AGTAAATAT	GGGTAAACGC	AGGTAAAATT	TTCTTTTCCT	2940
CTGTTTTTATT	TCCATAAAAT	TCAGAACGAA	GTCCCATGCT	GACCAAACCA	CCGAATCCAC	3000
CAaTATTATT	GATTGTATTG	ACAATCCAGC	TAGATTCAAG	GAGTTCCCGT	TTTGAGACTT	3060
TCAAGTTTAA	AATGCGATTT	AATGTTAAGT	CATAATTTAG	CATCGGAGTA	ACGGAAACTA	3120
AACCGATAAC	CATCATCAAT	AACAATTTCC	ATAGCGGAAT	CTCATCAAAA	ATCTGTTTTA	3180
ACTGTTCAAA	TGAGATCGTT	TTACCAATTG	ATAAAAGTTG	GGCAACAACA	ATAATTGAGA	3240
CCGCAATAAC	AAACAGTGTT	TTTAGTAATC	CTGTATGCGT	TTTGATCCAA	TGAATAATTT	3300
GTTTCATTTA	GACACCTTCA	TTTCCTTTAA	GAATTTTTTCG	ATTGTTTTAT	TAAAAATATC	3360
TGGTGAGTTT	TTGGCAAATA	AATGACCCTG	TTATTTTACC	AAGACCAATT	TGCCGGCAGG	3420
AATGGcATCA	GCAATCGCCC	GTGAATGGGA	AAGCGCAATT	AAATCTCGAC	TGCCAGTAAT	3480
AACAAGAGTA	GGACAGGGGA	TGTGTTCTAA	ATCTTGTTTT	GTCAGTCCAG	TATCTTCAAA	3540
TAACAGAGCC	ACGATAGGGA	GCCGTGCGCG	CAATTTTTTT	GAAAAAAGG	CGCCATACG	3600
ACAAAACCAA	TATTGAAAT	AAGTGGCAAT	CGTTACATAT	TTTTTCAGAC	CAGACATGGT	3660

GACATTTCCA GAATTTAAGA CGAGTTTGTT GACGAAATCA GGATATAAAT GGGTAAAAAC	3720
CATGGCAATA TTTGCTCCGT CACTAAAGCC CAATAAATTA ATTTTGTGA AGTGTTCCTG	3780
TTGAATAATA CCAGCTAAAT CTTCTGCCAA AAGCGGAAAA GTTAATTGAC TTTGTGTATT	3840
GGTGGAACGT CCATGTCGCG GGGTATCCAC CGCAATAACT TGATACTCCT TGGAGAAATA	3900
AGAAATTTGA TAATTAATAA AAGTATGATC ATTGCCATTA CCATGAAGTA ACACAAGGGG	3960
GTAATTTGTT KCTTTGCCAA AAATTTGAYa AAAAATTTGA CTGCCATCTA TTGTTTGAAA	4020
ATATTTTCT ATTGCCATTC TCCGTCACCT ATCTCATCTA TAATCAATCG AGTCCTCTCT	4080
CATTATAGAG TCTTTTGTCT TTTTTTGTGA ATAAAAACA GAAAATTTTC CGAACAGTTT	4140
TGTAAGGTGT TGTAATATAG TGTGATATAG TAGAAGATAG CCACCTAAAA ATTAGACTAG	4200
AGGAGGGAAA GTTGATGTCT TTAATCGTTT TAATTGGTGC CCAAGCTGTC GGGAAAATGA	4260
CCGTTGGCAA GGCTTTAGAA AAGCAGTTAG ATGCTAAATT ACTATTTAAT CATCAAACCA	4320
TTGACTTATT TGCAAACAT CTTGGGTATA CGGAACGTGC ATTTCAATTG TCTGATTCCG	4380
TTAGAAAGGA ACTTTTTCAT GCCTTTGTCT AAAATCCAGC CACAAATACA ACAAAAAACA	4440
TTATTTTAC AGTGGTGATT GCTTTTGATC AAGAACAGGA TTTTCAATTT TTAGAAGAAA	4500
TATCAGAAAT TTTTCTTAGT AAACATGAGT CCGTTTATTT CGTCGAACTC GTCTCTTCAA	4560
CATCAATCCG CCTAGAACGA AATGTTTCATC CAGAACGGCT AGAGGC AAAA CCTTCAAAAC	4620
GCGATGTCGA CTTTCTCGA AAAGAATTAC TCAATACCTT TAAAGAACAT CGATTAGTAA	4680
GCCACGAaGG TGAAC TAGAA AGTATATTTT CAAATGTAAA AATAGCGAAA ATAAATAATT	4740
CAAACATGAC ACCAGAAGAA GTAGCCACGA AGATAATTGC TACTTTTGGT TTGACATAAT	4800
GTGAAAGGAA CATCGGAATC AATGAAAAAG TTTTTTAAAT ATCTCGGTAG CTTGATGGTC	4860
CTTCTTTTAA GCATAGCTTT AATTTTGATA GGATATTTAA CCTTCCGAGA GTTTCGACCA	4920
CGAGCGGTG AACCAGTTGC TATAAATAAA CAATTTTCAG AGCAAACCAT CGAAAAAAG	4980
CAATCCATTT CTTTAGTCAC GTATAACATA GGCTATGCCG GCCTAGGACA AACAGAAGAT	5040
TTTTTTATGG ACGGTGGCAA AACTGTCCAA CCAGTTAATA AGGCAATGGT CCAACAGAAT	5100
TTAACAGGCA TTGAAGAAAC CTTAAAAGAA TTACCAGCGA TGATTTACTT ATTCCAAGAA	5160
GTCGATCGTC GCTCGCAACG TTCTTATGAA GTCAATCAAG AGGAAGAATT ACAAAAACAG	5220
CTACAGCTGA ACAGCGCATT TGCTTATAAT TTTAAGGTAG ATTATGTGCC CATTCCTTGG	5280
CCGCCGATTG GACGAGTAGA AAGTGGTTTA TTAACATTAT CAAATGAAAA AATAACAGAG	5340
GCCAAACGCA TTGCTTTACC CAATCCCTTT CGTTGGCCCG TTAGCATTAG TAACTTAAAA	5400
CGAGCACTTC TAGAAaCGCG CTTTCCAATT AAAGGCACAG ACAAAGAATT AGTGGTTTTT	5460
AACTTACATT TGGAAGCCTA TGACAATGGG GAAGGGAAGA TTGCTCAAAG CAAAAAGTTA	5520
GCTGACGTAT TATCCCAAGA ATACGCTAAA GGAAACTATG TAATTGCTGG AGGTGACTTT	5580
AATCaAGTTT TCaAAGGCAG TCATCGATAT CCTGACTTAG GGGAAGCTGG TTGGGTACCA	5640

GGAGAAATTG ATCCAGCAGA CTTACCTAAA CATTTTTCAC TTGCTTATGA TGATCAACAA 5700  
CCGACTGTGC GGGTGCCTAA TAAGCCTTAC ACTGGCTCTT ATGAAACATC CCAAGTGTAT 5760  
GTAATTGATG GTTTTATTGT TTCTGATAAT GTGGCTGTTT ATAGCGTGAA AACAAAGGAC 5820  
GAACAATTTA AATATACAGA TCACCAACCT GTGAAGATGG AAGTTGGGTT GAAATAGAAA 5880  
AAGGAGGAGA TAGATGAAAT TAGGGGTAT T GCCGCAATG GATTTAGAGC TTCAAAAATT 5940  
ATTGGAATAT TTTCCACCAC AACGAAAGAT ACAGCTAGCA AAAAATACAT TTTATATTTA 6000  
TGAACAAAAA ACATCGCAAG TCATTATGGT TTGTGCAGGA CAAGGAAAAA CGAATGCTAC 6060  
GCTATATAGT CAAATACTTA TTGATTCAAT TCAAATAGAA CAACTAATTA ATATCGGTAT 6120  
TTGTGGCTGT TTAAATGAAA AACTGCAACT CTTTGAGATG GTGCTAGGTG AGGAATACTG 6180  
CCATTACGAT ATTCGTGAAA GACAGTCAAT CAATAAATTT CTTATCAGC TGTACTATAA 6240  
AGGAGATAAA GGAATGCTGG CCGAACTTCA ACATCTAGAT GAAAAGATAA AGTGTGTTTCG 6300  
TTTTGGAAGT GGAGAAGGCT TCGTTTGTGA CACAAGTAA AAAAGAACT TAATTGAACA 6360  
ATTCTCTATT GACTGTGTGG ATATGGAATC CGCAGCTATT GCACAATGCT GCTTTTTAAA 6420  
TGATTGCGCT TTTGTTTCTA TAAGAATTAT TTGCGATAGA GCGGATGCTA ACGCAGTCAA 6480  
AGTAACAGAA GATACACAAA CACAGGCAAT GGAAAAAGTA TTTGAATTGG TCTGTGCATA 6540  
TATAAATTAA AAAAGAAAGA GCTTGAACA AAAATCCAGA AAGATTTTGT TTCCAAGCTT 6600  
AAAAAGTAAT ATAGGTTGGG GATAAAAGTA ACATTGTTAT TGTTCGAATT ATTTTACAGA 6660  
TAGCCTTTAG TAATAGAATG AGTGAATTT AGCATATTTG CTGGGGTTCC TGTA AAAAGA 6720  
AGCTGTCCGC CATTTTTTCC ACCTTCGGGA CCTAGTTCAA TCAACCAGTC GGCACTTTTA 6780  
ATCACATCAA TGTGATGTTT TAATAAAATC AAGCTGTTT CTTCTTCTAC CATTGATTG 6840  
AAAAGTTGAA GACTTTGTTG GATATCAATT AAGGTAAAC CATCTGTAGG TTCATCCATT 6900  
AAATAGATAG CTTTTTTTTG GTGAAGTGTG TCTGCTAATT TTACCCTTTG CAATTCGCCA 6960  
CCAGATAATG TTGAGAGCGA TTGATTGAGT TTAAGATAGC TTAAGCCAAC TTCCAGTAAA 7020  
TTTTTTAATG AAAGAGCAAA AGGCTGGTCT TTGAAAAAAT CATAGCCATC TTTAACGCTT 7080  
AAAGCTAGTA CTTCAACGAT ATTTTTCCTG TTATACAGAT AATGAAGCAC CTCTTCTTTG 7140  
TAACGCGTTC CGTGACAGGT TTCGCAAATA CTAGTAACAT CCTCCATAAA AGACATATCG 7200  
GAGACAATTA TGCCTTTCCC CTTGCAAGTA GGACAGGCGC CTTTGAATT ATAGCTAAAT 7260  
AAAGCTGGAC TAACATGATT TTCTTCGCA AACAGTTTGC GGACCTTATC AAAAATATTA 7320  
AGATAGGTCA TAGGTGTGGA TCGTAAATTT GCGGTAATGC TTTTTTGTGA AAGATGAATA 7380  
ATTTCTTGGT TATCAGCCTG GGCTTTTTGA TAAATTTCTT CCGCTAATGA TGATTTACCT 7440  
GAACCAGCGA CCCCAAAAT AACAGTCAAA ACTCCTAACG GAACTTCGAC AGATAGATTG 7500  
TTTAAATTAT GCAGTGTGCG ATGTTGATT GATAAAGACT TCCTTGCTTT TCTTGGTTGG 7560  
TCGTTTAAAG GGAGCGGCTC TTGAAGCGCT TGACTAGTCA AGGTTTTGGA GACTAAAAAT 7620

GCGTCATACG TTCCTGAAAA CTGGACATGG CCACCATTTT CGCCTGCGAA AGGTCCGATA	7680
TCGATGATAA AATCAGCTTC TCTAATTAAT TGTGGATTAT GTTCCACGAG AACACGGTG	7740
TTTCCTTTAT TTTTAAATT GAGCAATGCA CGACTGATCC GTTCGATGTC CTTTGGATGT	7800
AAGCCAGCAC TTGGTTCATC TAAATATAC ATAATATCAT TTAAGGCGCT GTTTACATAC	7860
TTAGCAATTT TAATCCGCTG TGCTTCACCA CCAGAAAGTG TTTCAGTTGC TCGTCCTAAT	7920
GTGAGGTAGG AGAGACCGAT ATTAATTAAT GCTTCTAGAC GTACAAGTAG CTCTTCTTGA	7980
ATAGTTTTTA TTAAGGATAG GTCCATTGAA CGAATAAAAG AATGTAATTC AGTGAGTGGC	8040
ATGTCAACAG CATCAGCAAT ACTTTTCTGA TTAATTTTGC AGCTACGAAC ACGTTCATTG	8100
ACTCTACTTC CTAAACAATC AGGACATCTT TTTACGGTAA CAAAGTGATT AAGGTATTTT	8160
TGATGACGTT TGCCTTCGTC TGTATGCAAT ATGCTACGTT GCATACGCGG GACGATTCCT	8220
TCATACAAAG CTGTATGAGG CCACTCTTTG GGTGGATTAG CTAGTTTTTG TTGTGGAGCa	8280
TATAAmmrTA ATGCTAACTC TTCAGGAGmw TAGTCCTTGA TTTTwtTATC CAAATCAAAT	8340
AGACCACTAT GGGCATAACG CTTCCACCGC CAGTTGC	8377

## (2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1102 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GTTTTTAAAC ACGAACGTTA AACGGCCGAA TGTTCCGTTT TTACGTTATT TTTGTAAATA	60
ACTAGGAGGC AGTAAGTATG AAAGTTTTTG ATTATGAAGA TGTCCAATTA ATTCCAATA	120
AATGTATCGT GAATAGTCGT TCTGAATGTG ATACTACAGT AACTTTAGGA AAACATTCGT	180
TtAAAATGCC TGTckTcCmG CGAACmTGcm AACatTATgr TGrAACgATT GCGGAacCCT	240
TGCTGAAAtG GTTATTTTTA TATCATGCAT CGGTTTGGAT GAAGAGGCGC GAGTGCCaTT	300
TATTAAAAAA ATGcAACAAA AGGGCTTGAT CACGTCAATT AGTGTGGGTG TGAAAGAAGG	360
GGAATACGCT TTTGTTGAAA CGTTAGCCCG AGAAGGTTTA GTTCCTGACT ATGTGACAAT	420
TGATATCGCC CATGGTCATT CAAATGCGGT GATTAATATG ATTCAACACT TGAAAAAATC	480
TTTACCAGAA ACGTTCGTTA TTGCTGGCAA TGTTGGAACA CCAGAAGCCG TTCGCGAATT	540
AGAAAACGCT GGAGCCGATG CCACCAAAGT GGGGATTGGA CCAGGAAAAG TTTGTATTAC	600
TAAGATTAAAG ACTGGCTTTG GGACAGGTGG CTGGCAGTTA GCGGCTTTAC GTTGGTGTGC	660
CAAAGCTGCT CGGAAACCAA TTATTGCAGA TGGTGGAATT CGGACCCATG GAGACATTGC	720
AAAATCTGTG CGTTTTGGGG CaACTATGGT CATGATTGGT TCaTTATTTG CTGGACATGA	780
AGAATCACCa GgTGrAACaA AAGTGGaAGA TGGTGTTGTT TACAAAGAAT ACTTTGGTAG	840

302

CGCCTCTGAA TTCCAAAAAG GCGAAAAGAn AAACGTTGAA GGCAAAAAAA TCTGGTTACG	900
CCATAAAGGG AAGTTGGCAG ATACTTTAGT CGAAATGCAA CAAGATTTAC AATCTTCCAT	960
TTCATACGCG GGAGGTCGCG ATTTAGAAGC GATTCGTAAA GTTGATTATG TGATTGTAAA	1020
AnATTCTATT TTTAATGGTG ATACCATTA ACATTCAACT ATAGACTGAA AAGAAGACCG	1080
CACGTAGnAA GGAAATTAAA AT	1102

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1009 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

nAAGAATACG GTCTGTCTTT AACAACTTCA ATTGATGATA TTTTAGAAGA TAAAGAGATT	60
CAAATCGTTG TGGAAATTAAT GGAACCAT TGAACAGCCA AAACGTATAT CATGAAAGCT	120
TTAGAAAAGG GCAAACATAT TGTGACTGCC AATAAGATT TACTGGCGCA ACATGGGAGT	180
GAATTAGTAG CGTTAcCCAA AAACATCATT GTGATTTATA TTACGAAGCA AGTGTGCTG	240
GCGGGATTCC AATTTTAAGA ACAATTGCTA ATAGCTTAGC GGCAGACAAT ATTCAACAAG	300
TTTTAGGGAT TGTGAACGGT ACGACGAATT ATATGCTTAC GCAAATGGTT TCTGCTGATA	360
AAAGTTATGA AGAAGCTTTA GCAGAAGCGC AAgcGTTGGG GTTGCTGAA GCGGATCCGA	420
CAAATGATGT CGATGGAATT GACGCAGCCT ATAAAATGGT TATTTTAAGT CAATTTGCTT	480
TTGGGATGAA TGTTTCATTA CCACAAGTTG ATATTCGTGG AATTCGCGGA TTGTCTTTAG	540
ATGATGTCGC TATGGCTAAA CAACTGGGGT ATGAAATCAA ATTGATTGGC TCTGCTGAAC	600
AAAATGAGAA TAGTATCTCG GTGGAAGTGG CGCCTATGCT AGTCAATCAA AAGCACCTTA	660
TTGCATCTGT TCGGAATGAG TACAATGCAG TATTTATTAA AAGTGCAGGC GTGGGGGAAT	720
CAATGTATTA CGGTCCAGGA GCTGGAGCTA AACCGACAGC AACCAAGTGTG GTCAGTGATT	780
TAATTACGAT TGCTAAAAAC ATTTCGCTTG CAACGACTGG CCATATGTTC AATTCGTATC	840
AACATAAAAC ACAATTGACG AGTTCCGAAA ATGTATTTGG GCAGTACTAT TTTyCATTAG	900
aTGTTCCAGA TACACCAGGA CaATTTTTCG AGTTGACGCA ATTAATGACT AAAGCAGAAG	960
TCAGCTTTGA TCAtTAGTTC AACAGaAATC AGATGGACAG CGAGCAAGA	1009

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5558 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

nCGGnTAAAA	TTTTTTCAGC	ATCCTTATAT	TACCAGTTTG	GTCTTTCAA <sub>g</sub>	AcTACTTGcG	60
ATcAGCaCtT	TgTAcAgTTA	aTTGTGCGCA	ATATGCCAaC	AaTCCAGaGC	TTTtAACTaG	120
TCAaTTATTT	GGACATCTTA	AAGgCGCTTT	TACGGGAGCA	GaTTCCGATA	AAATTGGGsC	180
CTTTCAAAGT	GCAGAAGGGG	GCGTTTTATT	TTTAGATGAA	GTTCATCGTT	TGAGTCCTGA	240
AGGGCAGGAA	AAATTGTTTA	CATTTTTAGA	TCAGGGTATT	ATTTATCGAA	TGGGTGAAAC	300
CAATCGGCCC	ATTCTGTTA	CGTGTGCGTT	GTGCTTTGCC	ACAACTGAAG	AAATTTCTAG	360
CACCTTTTTA	ACGACATTTT	TAAGACGAAT	TCCTATTCAA	ATTAAAATCC	CaAGTCTTGC	420
TGAACGCACT	CAAGCGGAAA	GAAAACAATT	AATTATGCGA	GCATTTTATG	AGGAACAGCA	480
AGCTATTCAA	AAAGCAGTTA	CTATCACGCC	ACAAGTTATC	CAACTTTTAG	AAAACCACCA	540
TTATGTAGGC	AATGTAGGTG	AATTAAGAAA	TAATATAAAA	ATTATTACTG	CACGCAGCTT	600
TGCCGCCAAT	CTTGATAAAT	GTGTTATTCC	AATCACTTTG	CACGATCTAC	CGAAAGAATT	660
TCTGGATCAA	TCAATCAAAC	TAGTTCCTGA	TGAAATGAA	TTGCCTATTC	GTTTAGATGA	720
GCAAACAAAT	CTGGTTAGTT	TGTTAGAGGA	AACAGAGCTT	GCGCAArGAA	GAATTATCCA	780
GAGTTATGAA	CGAATCCTTC	GTCTTTATGT	CTCTCACCAT	CATCATCTTT	CCACTGCTAA	840
TAACGATATC	TCGAAAGAAA	TTGAACGCTT	ATTGATGAT	TTATTATTTG	AAAAGAAGCG	900
AGAGAAAAAT	CACGAAATGC	TATTGTTTAT	TACGCAAAAT	ATCCGTCAAT	TATTGGAAAC	960
GATTGAATCG	TCTTATCAAA	TTCGGTTTAA	TGGCAGTTTA	GTCTATGCAT	TAAGCACCTA	1020
TCTGTTTCAA	CGAAGATGCA	TTGATTGGTT	TCCAGAACAA	GAACCTACAA	CAGTCATTGA	1080
CGAACTTTTA	ACAGAAGTTC	AAACCAAATT	AGCAACTAGT	TATGGtACGC	TGAACAACTG	1140
TTAACTTTAG	TGAAACGTTT	ATTAGATATT	GAACATCCC	AGATGGAwCG	AATTATTGTC	1200
ACCATTTATC	TTCATATTC	GGGTTTCACT	AAGAAAAGTC	ACTATCCGAA	AGCGGTGATT	1260
GTTGCCACAG	GATATGCCAC	AGCAAGTAGT	ATTGCGAATG	TTGCAAATCG	GCTGTTAAAT	1320
GTGCCAATTT	TTCAATCGTT	TGACATGCCG	TTAGATGTTA	CTCCCArAAA	AATTTCCGAA	1380
CACTTAATTC	ACTATATGGA	ACGTCAAGAA	ACGAGGAATG	GTTTAGTAAT	TCTTTTTGAT	1440
ATGGGTTCAT	TGAAAGAwAT	TTATCAGTAT	TTtCCGGCAG	AAGrAGAAGG	kCctTTTTTA	1500
tTAATGraTa	ATGtGACmAC	mAGcTTGGCT	CTGkCTaTwG	GAGrAGCCmT	TwAAGmCGAA	1560
kTTCTTTTGA	AGAATTgCCC	CAAAAAGcGT	TAmCGGTyCA	TCCmAATGaA	TGGGAAaTCa	1620
TATTACCTGA	AAATAAAACG	GAACgTGTCa	TTTTAACAAC	GTGTAgTACC	GGAATTGGCA	1680
CAGCCGTAA	AATTAGAGAT	TTACTTGAAA	AAAGTTTACC	TGCGGAAGCA	CAATTAAAGA	1740
TTATTCCTTG	TGAATATAAT	CAATTACGTA	ATGCAGAGTC	GATCAAAGAA	AGTTTTCCAG	1800
AATATGAGAT	TGTCGGCATC	ATCGGTACAA	ACAACCCGTC	TTCAAATGAC	CTCCCGTATA	1860
TCTCTTTAGA	AGAGTTGATT	GCAGGAAAAG	GAATTACGAC	CCTTTTAGAA	TGGACAAAGA	1920
GAGAATTAAC	TAAAGATATG	CTGTCCTATG	TAAATCATGA	GTTGATTTCGT	AaCTTTTCAT	1980

TAGATCGGGT CATTCAATCA GTGACGATTT TAGATACAGA AAAAATTATT CGACAAGTAG	2040
AGGTCTTTTT AATTCAATTA GAAGAGCGGT GGCAACAAAC AATTCAAAAT GATCGTAAAT	2100
TAGCGGTATA TGTCCATGTC AGTTGTTTAA TTGAACGATT AATCCGGAAT GAACCTATTG	2160
AAAATTATAA TGGCGCTGAA CAACTAAAGC AGTGTCAAAG AACGGTCTTG CAAGAGTTAA	2220
AAGAGGCGTT TAGTGTCAAT GAAAAAGTTT ATAGTGTCAA TATACCTGAA TCGGAGCTTT	2280
TTTATGTCTA TGATGTGTTG TTTGGAAAAA CGGAGTTTAA CAACGCAGAA AGCGATTTTT	2340
AAAATAAAAA GGAGGGAGTG CCACTTTTTT GGCACCTCTT TTGCTATATA AAGAAGTAAG	2400
TGAGGTGAAG AAAATGGAAC GTAAAATTAT TCTGGCAACA CATGGTAATT TCGCTGCAGG	2460
GATTTCAGACA TCATTAGCTT TAATTTGTGG TGAAACAAAG AATGTGGAAT GTCTTTGCGC	2520
TTATATGGAA CAACCTTATG ATTTAGCTAA GACGGTCAAA GACATTTTAA AGAACATCA	2580
ACAAAATGAA TTAATTGTGA TTACAGACCT ATTTGGCGGA AGCGTGAATA ACGAATTCTT	2640
TAATTATGTG GGGAAACAAG CAATGCATTT AATTCAGGA TTAAATTTAG CAACATTAAT	2700
TGAAATATAT ACCCAAATAA ATACTGTTGA TTCATTAGTT GATTTAGTAA AACGTGCCGT	2760
TGAAAACGGA CAAGAAAGTC TTTGTTATTG TAATGAGTTA TCTAGTAAAG AAATTTTAGA	2820
AGAAGaATTT TAGGAgGAat gAAnTGATTA AATTGTTAAG AGTCGATCAT CGCTTGTTAC	2880
ATGGTCAAGT GGCTTTTTCA TGGACCCAAG GCTTAGGAGC CGATTGTATT TTAATCGCAA	2940
ATGATGACGT ACCTAAAAAT GATATTCGAA AAACAACCAT TAAGCTAGCT AAACCAGCAG	3000
GAGTAAACT GGTCAATAAA ACGATTGAAG ATGCGATTGC CGCACTAAAA AGTGGTGTTA	3060
CAGAAAAATA CAACTTTTT ATTGTAGTTG AAAGTGTGGA AGATGCTTAT CGATTGGCTA	3120
CTGCTTATCC AGAAATAAAA GAAATTAAct TAGGAGGAAT GAAAGTAAGG GAAAATACTC	3180
GAAACATTTT TAAAGCAATT AATATCACAC CGGATGAAGA AAAAATGGTT AAAGAATTGG	3240
TCGCAAATGG TTGTGAAGTT GAAATTCGAC AAGTCCCAA TGACAAAAAA GTAGCAGCAT	3300
TAAATGTCTT GTAGAAAGGA GAAATAAAAA TGATAGGACA AGCGATTCTT TTAGGTTGCA	3360
TCGCTTTTAT TGCGCAGTCG GAATATGCTT TAGGAACCTC GTTGCTTTCA AGACCAATTG	3420
TAACCGGTTT ATTTGTCGGA ATTGTATTAG GTGATGTCAA AACGGGCGTA ATCATGGGCG	3480
CGACCTTAGA GTTAGCTTTC ATCGGTTCTT TTTCAGTGGG CGCATCAATC CCACCTGATG	3540
TAGTGACTGG TGGGATTCTT GGTACAGCGT TTGCTATTAC TGCTGGCGCA GGAACAGAAA	3600
CAGCCTTAGT CTTAGGCTTG CCAATCGCAA CGTTAACGTT GaTTCTTAAA AATATTTATT	3660
TGGGTTTGCT CATCCCAATC ATGAGTCAA AAGCTGATCA GTATGCTGAA GAAGGAGCAT	3720
ATAAAGGGGT GGAGCGCATG CATTTACTTG CAGGCTTTGG CTTATCGTTC aTGctAGCGC	3780
TCATTGTAAC TATTTCAATT GCCGTGGGCA GCAACACCAT TAAAGGCTTA CTTGATCTGA	3840
TCCCTGAATT TATTCAACAT GGATTGTCTG TTGCAACGGG GATTATTCCA GCGCTAGGAT	3900
TTGCGATGCT TGCACGCTTA CTGATTAACA AACAAGTGGC GCCATATTTT TTTTAGGTT	3960



TTGTCTTAAT GGCCTACTTT AAGTTACCTG TAACAGGAGT AGCAATTTTG GGTGCCATTA 4020  
 CAGCAGTAAT TGTTGTGAAT ATCATGAATT ATGCCAAAGG AAAGAATGAA ACAATACAAA 4080  
 CAACAAGTGG GGAGGTAATT GATGATGACG AAGACTTCTA AACTAAGAC AGAAACTTA 4140  
 ATTACAAAAA AAGAACTTAA TCAAGTCTTT TGGCGTTCAT TTCAAATGGA ATTTTCATGG 4200  
 AATTATGAAC GTCAGATGAA TATGGCTTAT GTCTATGCCA TGATCCCTAT TTTAAAGAAA 4260  
 TTATATCCTC AAAAAGAAGA GATGGCGGCA GCTTTAAAAC GGCATCTTGC TTTCTTTAAT 4320  
 ACTACGCCGC ATATCGTGAC GTTTATTTTA GGAATTAATG CTGCAATGGA AGAAGAAAAT 4380  
 GTTCTGGATG AAAATTTTGA AGTTGGAACA ATCGATAGTA TTA AACCTC GTTAATGGGA 4440  
 CCTTTAGCTG GTTTAGGTGA TTCCTTTTTC TGGGGAACCT TACGCTTGAT TGCAACAGGT 4500  
 GTTGGAACCT CGTTAGCTTT ACAAGGGAAT GTGTTAGGGC CAATTCTATT TTTACTCATT 4560  
 TTTAACGTCC CCCATGTGAT TGTTCTGTTAT TTA CTGAATC GTTGGGGCTA TAAATTAGGA 4620  
 ACTGGTTTTTC TCAAAAAGAT TCAAGCAAAT GGCATGATGG AAAGTCyCAC ATTAGGCGCT 4680  
 TCmATAATTG GCyTGATrGt GGkArGTGCy ATGaCCGCCT CCATGATTGA TTTAACTATC 4740  
 CCgATTACAA TCAGCgGAAG TGGGaAAAAT GCaGTCACGG TyCAAAATAT CyTTGACGAT 4800  
 ATCATGCCTA AATTATTACC GCTTGCATCT TTTGGTTTTG TTTTTTATTT ATTGnAAAAA 4860  
 GAAGTAAAAC CTTTAGCGAT TTTAGGTGGC ATGGCAATTG TCGGAATTCT TGGCTCATTA 4920  
 ATTTGGTCTAT TTTAGGAGGC AAAATATGAA AACGATGCTG GATTATrTAA ATGAGGAGCA 4980  
 aGCAGCTCTA rCTCAGaPTC TTAACCAATT TCAaTTAACT GATACAGACG TTTCAAAGGT 5040  
 CACGCACTGt CTAATTTTAG CAACTGGCTC GTCACATAAC GyTTkTCAGG CGGCTAAArT 5100  
 TTACCTTGAA GAAGTGGCGC CTATTTATGT GGAAATCCAA GAGCCGTTCA ATTTTGCTCA 5160  
 TTATGGCAAA GTGGATGATC GGTTTGATCT GGTCTAGCG GTTTCGCAAA GTGGTAAAG 5220  
 TGCTTCAACA ATTGGATGCC ATTGCAAAAA TTA AAAACCA AACGTCTGTA AAAACAGTGG 5280  
 CCTTGgACAA GTGaTGTCAC CAGCCCGATT ACGGaGGTTG TGGaTGaAGT CCTGGaTTTa 5340  
 AAAATAGGGc TTGAAACAGT TGGtTTTGTC ACAAAGGGt ATACGGcAAC ACTTTTaAAT 5400  
 TTAtTTCTAt TTGGtTTgAA ACTGGGCTAT CAAAAGCATC AATTAACAAA AGTAGAGGTT 5460  
 GAAGAAGAAT TAGCAAACT TnACCAAAGC AATTCAGGAA ATCGATTCCG TGATTATATA 5520  
 AACCGAACAA TTTTGGGCAA AAAACGGAAA AAAGGAAn 5558

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9047 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AnATnTGAAT CGTATAGCGA AGCAATTCTC TGATCAGAAT nATTTAACCTT CGTTTGGGAA

AAATTTCTGG CGCAAAACGT ATAACGGAAA ATCACCCACT CAGGAACTAC TGAAGGTTGC	120
AACGGAGGAT TTAGTACAAC ATAAAATGAT TAAACAGCTG GCAACGGAGT TAGAAATAGC	180
CCACTCACAA ACATTTGGGc AAGAAAAAAA AGAATGGCAA GACCAAAGA GCAGTTTAAC	240
GCTCTGGCAA TTTTGTAGTG CAAAAGATCA ACAATTACAG GACCAAATAA AAGAGAAATT	300
AATGGAGAAA GAAAAGCCAA CGCAAAAAGA ATTGCGGcAG GCCTTTGAGC AATTGGATGA	360
TAAATwTAAA AAGACCGATT ACTTTGTTGA AGCAATTGAA ATTCCAAATT TTTCAGGTAA	420
GCAAGCAGAA TTAGAAAAAA TAGCCGAAGC AATCTCTCCG AATCTATCTT ATGAGGAGAC	480
GTTATTGGAG TGGCAAAATA AATTACCTAA TCTTGTTATT GAATCGTATC AATTGAAGAG	540
TGCAGAAATC CAAAAGAGG ATATTTATTC GTTGTCTGTT GGGGAGATTC TTAGTGAAAA	600
AGCTGTTGGT ACCGTGGTGA AAGGATACCA CGAAAACCAA TACTTTTATA TTTTAAATAA	660
GGAAGGAGGG CACTTTTAC AGTTTGAAGA AGCACCACAA TTTGGCAAAA ATGCCTATAT	720
CAACACACGC TACAAAGAGA AaTTAGCTAC GTATCmAcmA GCAACCAAAG TAGACTTGTT	780
GGAGAAAGAC AGAGAAAAGT TTTTTCAAAA TTATCAAAAT AAAACAGTT AAGGAGCAGA	840
AAATGAGAAA AAAAGCGTAT TTTAAAAGTG TAACCATCTT CCTTCTATT TTGCTTGTTT	900
TTTCCCAGCT TCAATTATTT AGTCTTCCTG CTTATGGGGA AACAGTGAGT GATCAACCGT	960
TACTTTTTCG TTCAGTAGGT GTCGCTCAAA GTGGGACGAC CTACTATGTG GATGGGGAAG	1020
GAGGAAATAA TGCCAATGAT GGGCAGTCGC CAGCCAGTGC TTGGCGTGAT TTGAAAAAG	1080
TGAATCAGAC GGAGTTTCAG CCAGGCGACC ATGTTCTGCT GAATGCGCAA AGTACCTGGA	1140
ATAACCAATT GTTGCATCCA AAAGGGAATG GGACAGCGGC ACAAAGATT GTCATTGATT	1200
TTTATGATAC GAATGACAAA GGAGAAACCA TTTTGAAC AACACGGCGT CCGATTATTA	1260
ACGGTGGTGG CACCTACAGC ACAGGGACTT TCAAACGTGC AATCTCAGGG GCCGTGCAGT	1320
TAGTGAACCA AGAATATTGG GATATTTCAA ATTTAGAAGT GACAAATACA CCAGAGCTAG	1380
ATAATCTCGA AGGCTATAAG AAACCAGGAG ATGCTCAAAG AGCGGGGATA TTAGTACTAG	1440
GATATGAACA AAATCGAACG TTCAATAGTG TCACGATAAG GAATAACTAT GTTCATGATG	1500
TTCAAACAGA ATATTATTTA AATCTAAGTG GGAATACTGC AACCAAACGA TTAAGCGG	1560
TAGGTGGAAT TATTGTTTTA GGAAGTTGGT TTGATGAGAA TGGCAATGTA GTAACGGCAG	1620
CTAATGATCA TCGAACAACG ACTGGATTTA ATGATATTTT GATTGAAAAT AATGTGATTC	1680
AACGAGTGGG ACTTGAAGGA ATTCTGAACAA AAGCTGATTC AGATACcCA AGAGGCAACA	1740
CTTCTACAA AACGTTTCT AATATTACGA TTCGTAATAA CTATTTAGAA GATATTGCTG	1800
GCGACGGGAT TGTCTGTCA GAAGCCAAAA GTGGCGGTGT CGTCGAAGGA AATGTTGCTG	1860
TCCGAATGTG CAATGCCGAT TATGGTACAC AAAATTACGC GGGCGTTTGG GCAATGTCTG	1920
TAGATGACGG ACTGTTCCAA TATAACGAAg TATACGGAAT CAAATATGGA TTTAACGATG	1980
CTGAAGCTTA TGATGTGGAC ATGCAGTCAA ACAATGTTAT TTATCAATAT AACTATTCAC	2040

ATCATAATAC TGGCGGCTTT TTATTGCTTA TGTCTGATCA AAAAAATTCG GTTATTCGGT	2100
ACAATATCTC AGCTAATGAC GGCGGTGGTA ACCGAGGAAC AGGTAAAGAC AATCCAGGCG	2160
GAGCTGGTGG CTACAATTAT AAAGAACAAA GCATTTTCCA TTATTGGGTG AAAAAATGATG	2220
GAGCGGCAAT GCCAACTATC CACAATAATA CGATTTATGT TGGCGACGGT ATTTCAACAT	2280
CATTATTTGG TGAAGGAAaT AGTTCTGACA ACAGTGGTAC AGTAGCCAAT TTTTACAATA	2340
ATATTTTATA CAAAGAAGGT ACCGGACAAT TAAAATTCTT ATCTAATTAT CCAACAAATG	2400
GGACACAACC AATTGAACGT AAAATGGTCG ATAATCCAGA AAAATATTTT AAAAAATATG	2460
TTATTTGGCC AAAAGAAATT GCTACAGAAA AAAGTGGTGC AACTGTTGAA AAATTAGTAT	2520
CAAGTGGAAA CATTTTTGAA AAACCACAAT TAGAAATCAC AGATAACCCT GAAAAAGTGA	2580
AAGAGTTGGC AGAGCAAGAA TTTACGACAT TGAAACCAAC AAAAGATAAT GTGGTTGAAT	2640
TTACTTCTAA AGAACGCTTA CGTCAGCGAG CACAAATGTT CCGATTAAAA GAAAACTCTC	2700
CAGCGATTGG TAAAGGCTTG TCGGAAGTGA ACAGTCCGGC CGAAGACTTC TTTGGAAATA	2760
GCTTGAAAAA TAAAGTTTTA GATATTGGTG CGCAACAAGC ATCCACAATT GAAAAAGTA	2820
TTCGTTATCA AAATCAAGTG TTAGAAATTA GTTCGGCAAC AGGTGTGTAT CCTAATTGTC	2880
CAGAGCAAGT TGAATTAACG TACGAAGAAG TCGTTAATGA AGAAGTGGTT GCGACAGGCA	2940
AAAAAGAATT TAAGTTACAA TGGGAAGCAA TCCCTCAAGA AAAAATTAAT ACAGCAGGCA	3000
CAGTTGAAGT TGCAGCAACA GTTATCGGCT TACCGATAGA TGCTGTCAA GTAACAGCCA	3060
AGGTCTCTTT TGAGGGTGAA CTTGGTGAA GAAAAGACAC CGTTAAACTA AAAACAGCTC	3120
AAACTGCCTA TGTTCAAAAG AGTGATGGTA ACCGAGCATA CTCTGCTATA GCAGGCGGTA	3180
CCGCTGCCAT TAGTTCaGGA GACGCCTATA AATATCCTTA TGGAGTAAAC TATACGGGCA	3240
ATTACGCATT -AAAATTA AAAA AATGCTAGTT CTGCGGGGTA TAATCGCCGG ATTTACGTTG	3300
AAATTGATAC ACAAGAATTG AAGAACTATC AATCACTAAA AAGTGCCAAT TTAGAACTAA	3360
ATGTGATGCG TTATGATGCG TGGAATGGTG CTGGAATAC AAATGATGAA CGTTTGAAAA	3420
ATACGCAGTT CCAAGTAGAT GTTTATGGAA CAGATACTAA TTGGATGAGT AACACCATTA	3480
CTTGGAACAA TGGTCCAAAT AATTTAAATG TACCAAATGA AGAATTTATT GCTCGACAAT	3540
CTTTTACAAA TAGCTCAATT ATGAATAATC AAAATACAAT CTCTATTGAT ATTAGTAACT	3600
ATCTACGGAA GCTAATTCAA AGTGGTGAGA AAATCCAGC CAAGTTAAGT TTCTTATTAG	3660
CGATTACAGA CAGTCGTTTG CCTGGTTATG ATTCTGATAA TGCTGGGTTT GATGCTTTTT	3720
CAAAAGAAGG CGCACAAAAG GCCTATCAAG ACTTTTTTAAC AGGCAAATTG ACCTTGCCAA	3780
CTGGTCAACA GCTAGCAGAG GATAGTTTAG CTCCTAAAAT TGTTTTATCG AATGTTTTCG	3840
AAGTCAAACA TGAGTCGATT GAGGTCACGA CAGAAGCGGG TCAAGCCCCT AACTACCAG	3900
AAAAACAAC TATTTTTTAT TCGGATGGTT CACAACGAGA AGTGACTGTC AATTGGTCAG	3960
AAGTTCCAGC ATCTTCTTAT CAAAAGAAG GAATATTCAC TGTTGTAGGA CGAGCAGCAG	4020

GTGTCTCAT	SCCAATTATT	GCCAATGTAA	AGTCACCGC	TAAACATATT	GTTGGATTCA	4080
AGGAATTGCC	TGCSCTGGAT	CGACTAACAG	GCACTCGCG	AGGCGAAGT	AATTTGCCAA	4140
CTGAAGTAAT	TGCAAATTA	GATGATGGTA	GCGAACGAA	ATTGAAGGT	ATTTCTTGGG	4200
ATGACGATGT	TAGTAATTAT	TGCGCAAGTA	GCCCGCCAGG	TACTTACCAG	TTCCCTGCTG	4260
CAGTAGAAGA	AAAATAGGA	ATTGCGAATC	CTGATGAACG	AAAAATCTTC	CAAGTCGTGC	4320
AAACTCATGC	GATTCCGAA	AGAATTCAAT	TTGCTACAGA	AACAGCCACT	ATAAAATCAG	4380
GCGAAATTA	TCAATTCAA	TCTAAAGTGA	TTGGTCAAGC	GCCGCACACA	GAAACGGATG	4440
CTTGGAGCTC	GCAAGTAACG	TATGAGTTGG	TTACACCAGA	TGCTGGCAAT	ACAGTAAGTG	4500
TAGATGAAA	CGGTCTGATT	AGGACAGAAG	CAACGACTGC	TGCAGGAAAT	TATCAAATCA	4560
AAGTAACAAG	CAAAGTTTTG	CCGATAGTGA	CTGCACAATT	TTCAATTAAA	GTCACAAAAT	4620
AAACGAAGTG	TTTGAAGAGA	CTGGAACAAG	AATTGCTGTA	CCTTGCAATC	TTTCTAGTCT	4680
CTTCGATTAT	CTAGAATCAA	TAAATATTCT	ATTGCCTACT	ATTCATTCTT	TTTAAACAA	4740
GGTAAATGAA	TAAATGAAAG	TAAACGATAA	AGGTGTGTAG	ACAATGAATA	CTATCTTAAC	4800
AGGTGATCST	CCAAGTGGTA	AATTGCACCT	AGGACATTAC	GTAGGTTCCCT	TAAAAAACG	4860
AGTAGAATG	CAAGCAGACC	CTACAAATCA	ATTATTTGTC	ATGATTGCAG	ATTTACAAGC	4920
ATTGACAGAT	AATSCAAGAA	ATCCCGAATA	AGTTTCAGCC	AATGTTTATG	AAGTTGCTTT	4980
GGACTATTTA	GCASTTGGTT	TAGACCCTAC	GAAACACACA	ATTTTATCC	AATCGCAAAT	5040
TCCACASTTA	GCTGAATTAA	CAATGTACTA	TCTAAATTA	GTGACCACGT	CACGTGTTCC	5100
TCGGAAATCCA	ACGGTCAAAG	CAGAAATTGA	GCAAAAAAAA	TTTGGGGAAG	GTGTTCCGAC	5160
AGGATTCTTT	ATATATCCCG	TTTACACAGC	AGCTGATATT	ACTGCCTTTC	AAGCGAATTT	5220
AGTCCCGGTT	GGGGAAGACC	AAAACCAAT	GTTGGAAGAA	GCGCAAGAAA	TTGTGCATAG	5280
TTTTAATCAG	ACGTACGGTG	AAGTTTTAGT	TAGACCGGAA	GCTGTGTTAC	CACCAAAAGG	5340
CATGGGACGC	TTGCCAGGAA	TTGATGGCAA	CGGAAATATG	AGTAAGTCTT	TAGGCAATGG	5400
TATTTATCTT	TGAGATCCAG	CTGAAGTGGT	ACAGAAAAAA	GTAATGAGTA	TGTATACCGA	5460
TCCAAATCAT	ATTCTATAG	AGGACCCAGG	ACAAGTTGAA	GGAAATATGG	TCTTTACTTA	5520
CTTGGATGTT	TTTGGGAAA	ATAAAGATA	TATTGAAGAA	TTAAAGAGC	ACTATCGCCA	5580
TGGCGGTTTA	GGAGATGTGA	AAATTAAACG	TACTTAATT	GATGTCCTAG	AAGAAGAACT	5640
CGCACCGATT	CGTCGACGTC	GTGAAGGCT	GGCTAAAAAT	CCAGAAGCGA	TTATGGAAAT	5700
GTTGCATPAA	GGAGTCTTG	CTGCTGAAAA	AGTGGCAGCA	CAACCTTAA	CAGAAGTCAA	5760
AAAAGCAATG	GGATTCAAT	ATTTCTAAGC	ATAAAGAA	CTGGAACGAG	TCAAAAGAAG	5820
AATGACTTGT	TCCAGTTCTT	TTAAGTGGT	TGAAATGA	CTATTATATA	AGCTGGCGTA	5880
AAGAGTCGGA	GCTTGTAATA	GTTCCTGATG	CGTTCCTTTT	TCAATAATAT	CGCCATTTTT	5940
CATCACTAAA	ATTAAATCAG	CATTTTCAAT	TGTGGATAAT	CGATGAGCGA	TAACGAAGCT	6000

TGTTTCGATTT TCTGTGACAG TTTCCATCGC TTTTGAATA TGGGCTTCTG TTCGTGTGTC	6060
CACACTCGAA GTTGCTTCAT CGAGAATAAC AACGGGCGGA TTTGCTAAAA TGATTCGGGC	6120
GATGGTTAAT AACTGTTGTT GCCCTTGTGA TAATGCACCA TTTTCGCTAG AAATAATTGT	6180
GTCATATCCT TGAGGAAGGG TCCGAATAAA ATGATCACAT TGAGCAATTT TAGCTGCTTC	6240
AATTATTTCT TCACGAGAAG CATCTTTTTT TCCATAGGCA ATATTATCTG CTACGGTTCC	6300
TTCAAATAGC CAAGTGTTTT GTAAAACCAT GCCAAATAAA TTTCTTAGAT TTTGTCGAGA	6360
AAGTTTCGTA ATATCAATCC CATCAAAAGT AATGGCACCT TGATTTATTT CATAGAAACG	6420
CATTAATAAG TTGACTAATG TTGTTTTACC AGCGCCGGTG GGTCCACAA TGGCCACTGT	6480
TTTTTTCGGT TGAACAGAAA AATCAACATT CTTCAATAA ATTTTTTCTG GTGTGTAGCC	6540
AAATTGAACA TTTTAAATT CAATGGCTCC TTAGGAGAA GAAATAGTTT CTAAATGAGT	6600
TGCTTCTGGC TGTTTCATCAG CTTTCATCTAA GATAACAAAA ATCCGATCAA TGGAAGCCAT	6660
CGCTGCTTGA ATTGAGTTAA TGACATAAGA AGCGGTAGAA ATCGGCTCAG AAATTGGTT	6720
GATATATTGC AAATACGCTT GCAAGAAACC AATCGTAATA CCACCAGATA AACGAGCAT	6780
TGCGCCTAAG ATAGCACTAA TAATAAAAGC CAACTGATTA ATAAAACGAA TAGCTGGATA	6840
AATCGCAAAA TTTAGAACT GTGCTTTTTT AAAGGcACGA TAATGTTGTT GATTAACAGC	6900
ATCAATTGTT TTTTCAGCAT TTTGCTGTTG ATTAAGGTT TTTGTGACCA AATTCCTGC	6960
TAAATATTCT TCCATCTTAT TATTTAATTG ACCTAATTCA GCTTGACTTT GATCAGCAAA	7020
CACCTTGTTT TTATTCGCGA TTTTGTGCT CATGAAAGTA CTGCCGCCAA TTAGGAGCAA	7080
CACTAAAATG GTTAATTTTG CATCAATATA GAACAACATG ATTCCTGCAA AAAGGATAGT	7140
CACGACGGAA GTGAAAAATT GGTGATGCC TGTTAAAAGA ACTTGTGACA ATTGGTTTAA	7200
GCCAGTTGTT GAACGACTAA TGATATCGCC CACTTGGTGA TTGTCAAAGA AGGCCATTGG	7260
TAAAGTTTTA AACTTTTTTG TCACTTCTTT TCTAATCTT AAAGTAACCC GTTCGCTTAA	7320
AGAAGCCATT GCTCGTTCCT GAATAAAGA AGTGATACTA CTAATGATTG AAAAGAGAA	7380
TAAAAGTAAG ACGGGACCTA AGAGCGCTTC TTCAACTAAA GGAAGTGCA TACCTTTGAG	7440
CCCGACACGC TTGATAGCTT CTAAGAGGTT ATCAATCCCG ATTGCCATAA TATAGGTCAT	7500
GGCAACAATT AATAAGTTTC CAATTAACT GCAGATCATT AGACCATAAA AGATTGGATG	7560
CTCAGGTTTA ATCATTTTCC AGAAACGTTT AAAGCTTTGA AGATTCATTG TTTTTTTCAT	7620
TGTATGGCTT GAAAAGAAAA CTTTCTTCA AGCAGTCACA TCCTTCTTT AATTACTTGG	7680
GTAAGATACC TTGCGATTTA GCAAAGCTT GGTAATAGCT ATTAGTGGTA AGTAAATCAG	7740
CGTGGGTGCC TTGACCAACA ATCCTTCCTT CATCTAGGAC GATAATGTTG TCAGCGTTCA	7800
TGATTGTACT TAACCGTTGA GCAACAATGA GTAAAGTTT GTCCGACATT TGTGCATGTA	7860
AAGCGGCACG TAGAGCGGCA TCAGTTTTGT AATCTAATGC GGAAAAGCTA TCATCGAAAA	7920
TATAACGTC TGCCGGTTTG ATTAAGGCTC GTGCAATACA CATTCTTTGT TTTTGACCAC	7980

CAGAATAGTT GGACCCGCTT TGTGCTACGA AACTTTCAAT CCCTTGTGGT AAGGAATCGA	8040
TAAATTCAGA AGATTGTGAA ATTTCTAGTG CCGTTCTTAT TTCTTCTGTA GTTGCTTTGG	8100
CATTTCCCAT TAATAAGTTT GATAAGATTG TCCCACTGAA AAGAAAGGCT TTTGTGGCA	8160
CATAACTGAT GACTTGCGCA ATCGTTTGCT GAGATAATGA GCGGATATCT GTGCCAGAAT	8220
AGCTAATCGT GCCGGCTGTG ACCTCATTTA TTCGTAAAG TAACTTAACT AAAGTACTTT	8280
TACCAGCGCC AGTTGCACCA ACAATCGCCG TTGTTTTTCC TTTAGGAATG ACAAACCTAA	8340
CATTTTCTAA AACAGGATCA GCTGTTTCTG TGTATTGAAA CGTGACGTGA TCGAAGCAAA	8400
TCAAGGCTTT CTCAGGATGA TAAGCAATGG TTGCCTCAGT ATGAGGATCG CTAATTTCAA	8460
TTTCTTCTGA CAAACTTCT TGCAAGCGAT TCAAGGAAGC AACGGACCGT GGAATGACAA	8520
CTAATACCAT AGCCGCGATC ATTAAGTAGG CCAAAGTTAG AAGCGTATAT TCAATCACAG	8580
CTGTAACGGT GCCAATTTGT AGGGTGCCTA CTGCAACTAA ATTACGCCA AACCACAAAA	8640
TAGAGGCGTA CACAACGCCC ATTAATAAAA AGGCAACGGG AGTAATCCA <sub>g</sub> CAAAGATTG	8700
ATTCACTTTA CTCATGCGTT CAGCATAATT TTTAAATGTT TGATTGGTTC GCTGCTCTTC	8760
AAAATCTTGA TTATTGAACG CACGAATCAT ATTAATTCCA GTAAAGAACT CTCGTAAAGT	8820
CGTAGTAATT CGATCCATTT TTGGTTGAAT TTTTAAAGAT AAAGGGGTAC CTTTTTTCAT	8880
TAAAAAATAA ACAACGACTA AGTAAATAC GATGGAAGCC AATGTTACCC AAGCGAGTGA	8940
AACTGAATAA GAAATGTCA TCACTAAGGC AAATAAACTA ATAATAGGCG CTGGAAAGAT	9000
TAATTGGCAA AACAGCACAA TCATTGTTTG AACATTATCT ACATCGT	9047

## (2) INFORMATION FOR SEQ ID NO: 13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1751 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TTGTAATTTT ATTGGGTAA CGTTATTTT TGGTATCGG ATCTTAGTTG CGAATGGCGC	60
CAAAATTGTA GCGAATAAAG GGAAAAGAG TAAGGG <sub>g</sub> ACC ATTCTATTAA TGGTGATGGA	120
ATTGCTCTTG tTAGGTATCT TTTTATTTGT GaCTTTGCCa AGTAAAAATC TTTCaACCGT	180
GTTATGGATT AATCTGATTG GCGCAGTGAT TGCTAGTGGG GTGTTATTAA GTACACGCTT	240
AGCTAGTAAA GTCGAAGTTC CTGGTATCCA AAAGGGCCGC AAAGAAGTAG GGAAGAATAC	300
CAACAAAACG GCTAACAAAG ACTGGTCTGG CAAGATTTTG GGTGTTGCCG TTTTGGCAAT	360
GATTTTGTG <sub>g</sub> .ACTATTGTCA GTAGTATTAC CAGAATCTCT TCTATTGATG AAATTTATCA	420
AACGATTCCCT TTAAAAACGG AAGAAAAGC TGAAGTTTTA ACTTCGGCGA AAGAAACACC	480
CaTTGCGATT GCACCACAAA CGGCGAAACG AAAAATGTTA CAGAAATTTT CTGtCATTCC	540

CAATTcCAAT ATGTTTACTT TGGaTGGtAT AACGGCACAA GTTGTCAATG GCGAGTATGT	600
TTATGTAGcm ACAGTCGAAT TTAATGGCTy CTTTAAATGG CTAAAACTGG GGGAAgTACC	660
TGGCTATTTC ATCATTAGTG CGACAGATAT CAACGCACAG CCAGAATTTG TAGAAAAACC	720
AATTGTTTAT ACTCCTTCAG CTTACTTTGG CAAAGATGCC GCACGAAAAA TTTACTCCGC	780
TTATCCAGGT TATGCAGCAA CGGGCACAAAT CAATTTAGAG CTGGACGACC AAGGGAATCC	840
TTACTATATT CAAACGCTTT ACAAAGAGTA TGGGGTTTCT GGTCGCATGC ACTATAATGA	900
ATTTAAACG GCAGTATTGA ATGCTACAAC AGGAGAAGTT AACGTTTATG ACAGTCAAAA	960
AGCCCCAAGT TTCGTTGATG CGCCGATTAC TAGCTCAGCG GCGAATAGCT TAAATGAATT	1020
CTTTGGTCGT TACAGTCAAG GCTGGTGGAA CCAGACaATG TTTGGTGCCA AAAAAGATGT	1080
CAAAATTCCA ACTGAAAATG GTATTTATGC ATCAGGACAA ATTACGCCGA TGATGAACAA	1140
GGAAGGGAAC CAGTTGCTTT ATTTTACGGA CTTTACAAGT AGTGAAAAAG ACCAAGATTC	1200
CGCTTTAGGT TATTCCTTAA TCAATGCCCC AACAGGCCAA GTTACGTATT ATCGTGATAC	1260
TAAAGTCGGC ATTATGGATA GTGATGGCGC CATTTCTATC GCTTCAAAAA TTTATCCAGA	1320
AAAGAAATGG AAAGCTAGTA TGCCAGTCCT TTATAACATT GATGGCGTGC CAACGTGGAT	1380
TGTTTCGTTA ATGGACAGTA AAGGGATTTT CAAAAAATAT GTCTATATAA ACGCCGTAGA	1440
TAATGATATT GTAATTGATG CAGACACGGC ACAAGGCGCA TTAGACGCTT ATCGAATTGA	1500
GTTAGTAACA AAAGGAAGCA ACAATGAAAA TACAGACAAA GCCAACTTGA AAACGATGCA	1560
AGGTAAAGTA GCTCGTGTCT CAGTAGTAGC TGGGGAAGCG CAAACGGTAG TTTCATTTTT	1620
ACTAGAAGAT GAAAAaCCA TTTTAGCGT AACGaCTAaT AACAGTcCcT GGCGCTATTC	1680
TTAAAaGAGG GCGaCCAAGT GACaTTTAAA GCCGTAGtGA CTGAAAATGC TAAAGCAGCG	1740
AATGTTGAAG A	1751

## (2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3928 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CGTTTATTTT ATCAAACCA AATGGATTG TAAAACATT CACATGTTAT TTAATAAACG	60
ACTATTACTA GAAAAAAGAC AAGCCAAGTT TCCTTTATTT AGGAAACTTG GCTTGTCTTT	120
TTATCTTTTT TACTTCGTTA TTTTGTGATA GTTTGCGAAT AATTGATTCG CATCTCCTTC	180
TCCATCAATT TTCTTGAGAT GGCCGTTATC TATTTTAAAC ATAGATGGCA CATGTAAAAC	240
TTTATTTTCT CGGATAAATG TTGACCATT CTTTCATGGT TCtTTGTTGG TTAATTCTAC	300
ATAGTAAACC TGTTTAGGTG CTTGTAATAG TTTCTCGAAA TCCTGACAAT CTGAACAGTT	360
TtTCTTTTTT AAAACAACAT AAAATGTTTC GCCAGAGTGT TTTTTTCTT GATACTCATT	420

GAAACTTAAT TCCTTTTTAT GCGGAATAAA TGTA AACCT ACCAACAAAC TAAGTAACAA	480
AATTAACAGA CTGAACGAAA TTTTTTATT CATGTTTTTT CTTCCTTTTT AGTTTGTCTA	540
ATTTTTCTT TTTTTATTTT TTGTTATGCC AAAGCTAATA GCCACTATAA GGACAATTAA	600
ACCTGCGATC ATGTAAAGTG GATTTACTGT TGTATTAGTC ATTGGTAAAT GAATTCCTTT	660
ACTTGCTGaT TTTCCAAAAT TGaCTGGGT CGTTGGTGCT TGATGATTTG TAGGCGTCAA	720
CGGTTTTTTC GGTTTTAGCG GTTGT TAGG TTCAGGTTTT TTTGGGTCTT CTGGCGTTGG	780
TGCAGGCGGT GTTACGGTTG GCTTGTTGGA ATGCAACACG TCACCTTCGT TGCCAAAGTT	840
TAAGTCGGCT TGGTTGGGAA TCCCGCCTTG TTCAATATAA GGCGCTAATT CTTCATCCGT	900
TGCGTCAGTT TTAATTTTAG TGGTAATAGT CATTGTGTAT GTATGACCAG CTAAGTAAGA	960
GTAGCTGTCA TCTTTTTTGT TCATAGTAAA AGTTACTTTG TTATTTTCTT GTGTTACTAT	1020
GCCATTATCT GTAACATCTT TGCCATTTTC mTCAGwAACT TTCACGTCTG TGATGTCTAA	1080
CACTTTATTA ATGTCATCCA CCATGCTGGC TtGGGTCCAT GTGCTTGTTT CGTTACCGAA	1140
AGCTGTTTTtG ACGTTCCATT TAAATTCTTG GTCACGGTTG GTTAAATCTA AATGTTCTTG	1200
GCCTTCGATA TCTTTTGTA TCGTTGGATC TTCTGGcGTT GGTGCAGGTG GTGTTACGGT	1260
TGGTTTGTTG GAATGCAACA CGTCACCTTC GTTGCCAAAG TTCAAGTCGG CTTGGTTGGG	1320
AATGCCACCT TGTTCAATAT AAGGTGCTAA TTCTTCGTCC GTTGCGCTAg CTTTGATTTT	1380
AGTAGTAATG GTCATTGTGT ACGTATGACC ACTTAAATAG TCATAGCTGT CkGctTGkTT	1440
GtTCATTTCA AAAGTTACTT TGTTATTTTC TTGTGTTACT TTGCCGTTAG CTGTAACATC	1500
TTTACCATTT TCATCTGTGA CTTTACATC AGTGATGTCT AACACTTTAT TAATGTCATC	1560
TACCATGCTG GCTTGGGTCC AAGTGCTTGT TTCGTTACCG AAAGCTGTTT TGACGTTCCA	1620
TTTAAATTCT TGATCGCGGT TGGTTAAATC TAAATGTCTT TGTCTTCTA CGTCTTTAGC	1680
AATATTTGGA TCAACTGGCG GTGGTGTTAC GGTGGTTTG TTGGAATGTA ACACGTCACC	1740
TTCAATTGCCA AAGTTTAAGT CGGCTTGGTT GGGAAATCCCG CCTTGTTCAA TGTAAGGCGC	1800
TAATTCCTTCG TCCGTGCGT CAGTTTAAAT TTTAGTGGTG ATAGTCATTG TATACGTATG	1860
ACCACTTAAA TAGTCATAGC TGTCTGCTTG TTTGTTCAAT TCAAAAGTTA CTTTGTTATT	1920
TTCTTGTTGTT ACTGTGCCGT TAGCTGTAAC ATCTTTACCA TTTTCGTCGG TGACTTTCAC	1980
ATCAATGATA TCTAGCACTT TATTAATGTC ATCCACCATG CTGGCTTGGG TCCAAGTACT	2040
GGTTTCGTTG CCAAAGGCTG TTTTGACATG CCAATCGAAA CTATCTTCAC GATTGGTTAA	2100
ATCTAAGTGT TCTTGATTTT CGATATCTTT TGTAATCGTT GGATCTTCTG GCGTCACGTC	2160
ACACGCATGC GCATTGTCAT TTTTAGGAAT CGTCGTTCTT GGAAATTCCT TAGCAATTGT	2220
TCCTTGATTC ATTGTCCAC TACTTACAAG GGTGCTGCA TCAATCGCTG TATTTTCTTT	2280
GACTTCGTAG TGAATCGTGA TGTTTCCTAC GTAACCTTCT GTACTAGTTG CTGAAATGGT	2340
TTGTCCGTTA ATTTGAACAG GCACATCATT TCCAGCATCG TCCGTAGCGG TCGCAGATTG	2400



AATATCAAAT TGATTGGCAA TCGTTAATGA AGCTGTTGCT GGTGTCGATT GCGCCAGACG	2460
ATCTTTGACA ATTTGTTTTA ATTGGGTTGT AAAATCATCA ATAGATTGGC TTGTAATAAA	2520
ATCTTCTGGT GTGCTAGACC CTTGTTGCAA CTCTTGTTTT ACAAAAGGAC CCACTTCTGT	2580
TTTATATTTA TCAAAGTATG AATTCACCTGA ACTTAAAGAT TCAACACTTT CCCAATACGC	2640
ATTAATCATT TCATAGCCTT GGTTAGTAAT TTCTTGTTTT AACGCTAAAA CTTCTGCTGC	2700
TGCACCTTGG TAGTCATTAC TATATTCCAC TGAGACTTGA AGAGGATGTC TTGGATCTGG	2760
ATATTCATTG ATTGAATCAT TGGTATTGGT CTTATGCAAG TAACCATCTA AACGTGTATT	2820
AGCGACCCCA TCTGTCACTA ATAGGAAATA CGTTTTTCGA TTCGTTAAAT CTCCGTGTGT	2880
TTGATTGTAC GTATCTAAAG CGAGTTTCAA TCCTGGGGCG GTTGGCGTAC CACCATACGT	2940
CCGAACGTCT CCAAAACCAG AGACAAATTG GCTTTTATCA TAAGTCAATT GCGTATTGAC	3000
GCGCACATTC ATATCATAAT CAGCTGAATT AATTTTTGTC TTTCCATCAG GAAACATAAA	3060
TTGTTTTCCG CCGCGATATG AAGCCAGCAT CACGCGGTCT TGATCAGATA AGCCTTGAAC	3120
CACTTCATCA ATCGCTTGTC TTACATGTGG AAAATTATCT GAAAACTAC CACTAGCATC	3180
TTCAACAACCT ACTAAATCGA CTGGTTCCGT TTCACCCGCT TGAACGGGA AAGTAGCTTG	3240
TCCTTCCAAT ATCCGTTTAC AATCAGCCAG ACTGTCGTCT TTCACTTTCA CCGTATTGGT	3300
CATTCCTTCT GTATCTCCTG CTTTAACCGC TTCTTCAGCT AAAGCCAATG TGGGACTGAA	3360
ATTTCCACA ACTGTTCCAG CGACAATCAC CATGTGCGAT AATAATTTCTG TTGCTTCTT	3420
AAACATTTTA ATCCTCCCTA TTTTCTTTCTG CTAGTGTTTG GGACACTTGC TTCTCTTTAC	3480
TTCAACTACA GTCCTTAAAT TCTTTCATCA AGTCGATGTG TCTCCTTGTT TTTGTGACAC	3540
ATCTCTTACT CAATTTCTGA TTCTATCACC TTTAAAGGCT AAAACTTGGC AGTATTCGGC	3600
TCTTGAAAAC GTAAAATTTG TCTTTTAAAA ATTAATGAAA AGACTATCCA CTGACAAAGT	3660
TGGTTTTTAC TTAATTTTCTG CTAATTTGATT CCTTTTAAGA CTCTTTTTTT ACACAAAAAA	3720
AGTAAAGAA CCTTTACAGT CCTTTTACTT AATTGTTAGA TGCTATTCAT AATTTTCTTA	3780
GTTTTGTTCT ATAAGCTGAA TAGTTTAGCC TGTTTATTTT AACAATATTT GACGTTAGTC	3840
AGCCGTTTCG TTATAGTTAA AGAAAAGGCT TTTTGGTAGG AGTGAATACG ATGAATGAGT	3900
GGCCCCATCA GTTAATGACG CTTTTAAC	3928

## (2) INFORMATION FOR SEQ ID NO: 15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5259 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CGTTAGGTAC ATTGCTGGTT AATGGAACAA GTGAACAGTT ATATGAAACA AaCGAmaCCA 60

TTCACTTGGC	TTTTCCAGCT	GAGAAATTAG	TTATTTTAGA	AAAATAAAAA	CCATTTTGGG	120
GGAAGATTTA	AAAATGAAAA	AACGCGTAAT	TTTAGGGACA	TTAGTCGCTG	CAACGTTATT	180
AATGACTGCT	TGTGGAAACA	GCGAAgcAAC	TACGAAAAGC	GAGAGCAAAG	GTGGAAGTAA	240
TGCTTTAGTC	GTTTCAACTT	TCGGATTAAG	TGAAGATATT	GTCAAAAAAG	ACATTATCGC	300
TCCATTTGAA	AAAGAGAATG	AAGCGAAAGT	TACCTTAGAA	GTAGGCAATA	GCGCAGACCG	360
CTTTACGAAA	TTAAAAAATA	ATCCCAATGC	GGGAATTGAT	GTCATTGAAT	TAGCACAAAGC	420
AAATGCAGCA	CAAGGTGGAA	AAGATGGGTT	ATTTGAAAAA	ATTACAGAAA	AAGAAGTACC	480
TAATTTAAGT	CAGTTAACGC	CGGGAGCAAA	AGAGGTTTTT	GAAAGTGGTG	CTGGCGTACC	540
AATCGCTGTA	AACAGTATCG	GGATTGTTTA	CAACAAAGAA	AAATTAGGCA	AAGAAATTAA	600
AAACTGGGAT	GACTTATGGT	CAGCTGATTT	GAAAGGTAAA	ATTTCTGTTC	CAGACGTTGC	660
CACGACGGCA	GGTCCTTTAA	TGTTATACGT	TGCTAGTGaA	CATGCTGGTC	AAGATATTAC	720
AAAAGATAAC	GGGAAGGCCG	CTTTTGAAGC	GATGAAAGAA	TTAAAACCAA	ACGTTGTTAA	780
AACGTATTCA	AAATCGTCAG	ACTTAGCyAA	TATGTTCCAA	TCTGGTGAAA	TTGAAGCAGC	840
TGTGGTTGCT	GATTTTGCGG	TTGATATTAT	TCAAGGCGCA	CAGAAAACGT	GACGTATATC	900
GTTCCAGAAA	GCGGCACGTA	TGCCAACTAT	AATACAGTCA	ATATTCCTAA	AAAAGCACAA	960
AACAAAGAAA	CCGCCCTTAA	ATTTGTTAAC	GCACGGATTA	GTGAAGAATC	ACAAAAGGCC	1020
AAAGCGATTT	CGTTAAATGA	AGGACCTACA	AATCAACAAG	TAACGTTAAG	CGAAAAAGAA	1080
GCGAAAAATA	AAACATATGG	TGCCATTGCC	GAACGTGCCA	AAACAGTCGA	TTTTAATTTT	1140
ATTAATAGTC	AACTAGCGGA	TTGGATTGAT	CAATGGAATC	GGACGATGAA	TCAATAGTTT	1200
TAGAAAAGAG	GGACAAAATG	AACGTCGATG	TATTACTAAA	AAATGTGTGG	CTGTATCAAA	1260
CAGTCACTCA	AACATTCGTT	CAACGAAATG	TTGCGATTAA	AAATGATAAA	TTTTATTACA	1320
TTTATGAAGA	AGAGAACGTC	AATTTGCAGC	CCCAAAAAAC	AATTAATGCC	GAGAATCAAT	1380
GGATGATTCC	CGGCTTAATT	GATGCACACA	TGCATATTGA	AAGTTCCATG	ACGACACCTA	1440
CTATTTTTTC	CAAAGCAGTA	GTGCGTTAcG	GTGTGACAAC	GGTGATTGCC	GATGCCCATG	1500
AAATGGCTFA	CGTATTTGGC	TTAGAAGGCT	TAAAAGCTTT	TATGGCCGCT	GAAACAGAAC	1560
TAGATATCTT	TCACGCAATT	CCTTCCTCGG	TTCCTTCGAC	CACTCCCGAA	TTAGAAACAA	1620
CTGGTGGGAT	TATCGGCTTA	GCAGAAGTAG	CAGAATTACT	CAAAGAGCCA	AAAGTGATTT	1680
GTTTAGGGGA	AGCCATGAAT	TTTAAAGGGA	TTTCCTATGA	GCCAGATTCT	TTAATTCGTC	1740
AAATTATCGA	TTTGTGTCAA	AAACAACGAC	CAACCATGCC	TTTAGAAGGC	CACTGTCCTA	1800
AAATCGAAGA	TCAAGAGTTA	GCAGATTTTC	TATATAGTGG	TATCACTTCA	GATCATACCC	1860
ATCAATTTCC	TAAAACGTTA	AAAGAAAAAA	TTGAAGCGGG	CGTCTTTATT	CAATTTCAAA	1920
ATAAATCAAT	CACTCCTGAA	AATATTCAGG	TGATAGTAGA	CAATGACTTC	TACAATTATG	1980
CAAGTATTAT	CACGGATGAC	GTGATGGCTG	ATGATTTGCT	AAAAGGTCAT	CTAAATGAAA	2040

ATGTCAAAAA	AGCTGTTTCAT	GCAGGGTTGC	CTATCGAAAA	AGCCATTTAT	ATGGCAACTT	2100
ATACGCCAGC	CAAACGAATG	GGTCTCCATG	ATCGCGGCGA	AATTGCACCT	GGTAAAAAAG	2160
CTGACTTTTT	GTTATTAAAT	GACTTAGAAA	GTTTTGATAT	TAATACGGTT	TATAAACTCTG	2220
GAAAGGTCGT	TTTTGAAAAA	GGCGAGCCCT	TCCATTATCC	AGAAAAATC	GAAGAGTTTC	2280
CAGCCACTTA	TCAACAAACC	ATCCAATGTA	AAAAGTTAAC	AGAAGAAGAT	TTACTTTTAA	2340
AGGTGGCTAC	AACAAAAGAA	ACAGTTCGTT	GTAATGTGAT	TCAAAAACAA	GAAATTGGCA	2400
CATTTACAGA	GCGAATTACA	AAAGAAATTC	CTGTTGAAAA	TGGATTGTTG	CAATGGCAAA	2460
AAGCCAACTG	CGCCTTATTA	ATTGTGATGG	AACGTTACGG	AAAAATGGT	AATATTTCTT	2520
TTTCTTTAAT	GGATCAACCT	CTTCCGAGA	AAGGAGCCAT	TGCGACAACC	TGGGCCCATG	2580
ATCATCATAA	TTAATGGTG	ATGGGTAATA	CGATTGAGGA	TATTCTGTTA	GCGCAGAATG	2640
AATTATTAGC	GATGCAAGGC	GGCTATTTGG	TTGCTTCGGA	TCAGCAAGTG	ATGGCAACTT	2700
GTCCTTTACC	AATTGGTGGG	ATTTTATCGC	AAGCACCGAT	TGAACAATTA	GGCGCTTCTT	2760
TACAAAAAGT	CCGACAAGCT	ATGCAAGCGT	TAGGCTATCA	AAATATGAAT	GAAATCATGT	2820
CCTTTTCTAC	ACTATCATTA	CCTGTCTCGC	CAGCCATTAA	AGTGACAGAT	TTTGGGATGA	2880
TGGATACGAA	GTCACAACGT	TTCTATCCAT	TAGTTTTTCC	AGAGGATGGA	GTTCTTTTGC	2940
ATGAAAACAC	TCATTAAAAA	TGTGCATATA	CTAACAATGG	ATGAGCAATT	TTCAGAAATC	3000
AAAGCCGGCT	ATTTAGTGAT	TGAAGAAGAC	ACAATTGTGG	AATTAGCACC	AATGACCACA	3060
CTTGATGAAA	AGCGAATGGC	TGCAAAATCA	GTAATCGATG	GTCAAAATGG	AATTTTAATG	3120
CCTGGGATGA	TTAACACCCA	TACCCATGTA	GGCATGATTC	CGTTTCGTTC	GTTAGGAGAC	3180
GATGTGCCAG	ATCGACTCCG	GCGTTTTCTT	TTTCCATTAG	AACAATTCAT	GACAAAAGAA	3240
TTAGTAGGAT	GCAGTAGTGA	TTATGCAATT	GCCGAAATGT	TACTGAGTGG	TATTACGAGC	3300
TTTTGTGATA	TGTATTATTT	TGAAGATGAA	ATTGCTAAAA	GTTGTGAAAA	AATGAGTGTT	3360
CGTGCTTTGC	TCGGAGAGAC	GATCATTGAT	ATGCCCACTT	GTGATAGTCC	GGAGCCTTCA	3420
GGCGGtCTTT	TTTACGCGGA	AACCTTTATT	CGCAAGTGGC	AAGGCCATCC	GTTGATTACG	3480
CCTATGCTTG	CGCCACATGC	ACCGAATACC	AATTCACCAG	AAGTGTGGC	GAAAATTATT	3540
GAACCTTAGTC	GGCAATACCA	AGTTCCTGTG	ACCATGCACG	TTGCTGAAAT	GACTTATGAA	3600
ATGGCTGAGT	TTGAAAAAGC	CTATCAAAAA	ACACCAATTG	CTTCTTAGA	AGAACTGGGT	3660
TATTTGAGCG	AGCCGTTTAT	TTAGCGCAT	TGTATTTTGG	CAACAGATGA	AGATCTTGCG	3720
AGTTTAGCTG	CTACTAATGG	AAAAGCGCGT	GTCGCTCATT	GTATCGGTGC	GAATACTAA	3780
TCAGCCAAAG	GCGTAgcGCC	GATTAAGCAA	ATGCTTGATC	AAGGGATTAT	TGTCGGTTTA	3840
GGCACGGATG	GACCTAGTAG	TGGGAATACA	TTAGATTAT	TCACCCAAAT	GCGCATGGTT	3900
GCGAATTTTC	ATAAGACAGC	ACACCAAGAT	CGCTCCTTGT	TTCTGCTAA	AGAAATTGTT	3960
TATCTGGCAA	CGATGGGGGG	CGCTAAAACG	TTAGGCTTGG	CGGAGCAAGT	CGGCTCATTG	4020

316

GAAGTGGACA	AAAAAGCGGA	TATAACATTA	ATTGAAACGC	AATCAGTTAA	TATGTTTCCG	4080
ATTTTGTATG	CCTATTCAGC	GTTGGTTTAT	TCAGCAAATG	CTAGCAATGT	TGAAGCCGTT	4140
TGGGTGAACG	GTCAACAGTT	AGTTGCCAAT	AAAGAATTAC	AACAAGCTAA	TCTCAAAGAA	4200
ATCAAGGaAA	AATTATATCA	GGCCATGAAT	ACGTTTGTGA	AAGAAGCTAA	AAAAAGAGCT	4260
GCTCTCTAAA	CAGAACAAGT	AGGACGTATC	GGATTGATTA	AATAACGCTG	GATTGACGTT	4320
TAGTGATTCA	GGTCTAATA	GAGCGCAGAC	AAACTTTTAA	TGAAAGAGTT	GTCTGTGTTC	4380
TTTTTTGTTA	TAAACATTTT	TCTTTTCTA	TATATATTTT	TAAGGTAAGT	TGTATTATAG	4440
TAAGTTTGAG	AAGATGAGAA	TGAGAGGAAA	AGCAATGAAT	ATACAACGGT	TTATAGAAAA	4500
ACGCAAAGCA	CGTGGTTTGT	CGCAAAGTGA	ATTAGCTAAA	GGGATTTGTA	CGCAAGTTAC	4560
GGTTAGTCGC	TTTGAAAAGA	ATGGTCAAGT	CCCAACTTTA	AAAATATTGA	TCCaACTATG	4620
TAATCGTTTA	GAGCTACCTT	TGGGGGAACT	CTTCCCGCGC	GTAGGGATTA	AACAACCTGA	4680
AATTCTTGAA	AAAATGGAAG	AAGCAGAATT	CTTCTTGATT	ACCAGCGAAC	ATGACCAATT	4740
GCAAACAATT	TTAAAAAATA	TTCCTTTTGA	TGAAATCAAA	GATTCCCAAC	TGTTATTAGA	4800
ATATTATTAT	TTACAAGGCT	TTGTCATGAT	TTTTCAAAAT	GCTTCGTTGA	TGGATTGCTT	4860
ATTTACCTTT	GAAAAACTTC	TGTTTGAAGA	GCAAAAATAC	ACGAGCGATA	TTTACCGTTT	4920
GTTAGCTTTT	ACTGGAATTG	GCATGGCCTA	TGCCAAAGAA	GGCGAAATTG	AGAAAGCCGA	4980
ATTTTATTTT	AATAAGGTCT	TCAAGGAAAT	TTATCTCTAT	ACTATTCAAT	CAATGGAAGA	5040
TACTTGGCGA	GTATTAAATG	TTGTCTTTCA	CTGTGGTGTG	TTTTATGCAG	AAAAAGGCGA	5100
TTTAGAAACA	AGTGATGCTT	TGTTAGAGTA	TGCGATTTC	ATCTGTTCTG	ATAATCACGT	5160
CACGTATTAC	TTGGCAAGAG	CAGCTTTCCA	ACTTGCGAAA	AATGCCTTAG	CTGAAGAAAA	5220
GCCGCAAGAG	CAAATTTTAG	AACTACTACA	AGATGCCCC			5259

## (2) INFORMATION FOR SEQ ID NO: 16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7684 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGGCCGTnTG	AAGTGGaCCc	TnGAgGAaTT	wGtTCaAAAr	GCmAAGnTT	TAATTGGaGA	60
AGGCcAmCGT	GGAACnTGCG	CaACmATTcA	TTGrAGAACA	TAAAGaTGAA	TTAGGCGAAC	120
ACTATCATAA	AATTACAGCG	ATGTTAGCTG	ATACAACgAc	CgATGgTTTG	TTTGATAAAA	180
TTAAAGGATT	GTTTCAATAG	TTGAAAAAAA	TCTGAGATAA	AAGTGTTTAG	CTCCAAAAAC	240
CAAGTAGGTA	CGATGATTTC	ATTTGTTTCAT	GCTTCACGGT	GAAACATTGC	AACTCTCGCC	300
GTTTAACAGT	TTTTGAGCGT	GGAGCAAAAA	TCCAAAGTGA	TTTTTGTCCT	ACGCTCTTCT	360
TTTTTTATTT	AACTTTTAAC	CATTGTTTTG	GCAACACATA	CGCAATTGGA	TTAAAAAGAT	420

AATCCGCTCC	CCGTTTCACC	ACAGGCATAT	TCGTAATAGC	TAAAATGAAG	AGGGCTAAGA	480
AGAAAAGTGC	CAGGATTCGG	AAAGTAAACC	ATTCTTTAAA	GAGTGTGTCT	AATTGCAAAC	540
CACTTTCCAA	CAGAATAATT	AAGGGCATAT	GCAAGAGATA	AGGGTAGAGG	GTATCTTTTC	600
CGAGGGCGCT	TAAAGGTAAC	TTTCGTTTCG	GCAATAATGC	AAACCAAGAA	AGAGACACCA	660
ATGCTGCAGC	AACGTAAAAG	AACAGCTGAA	TGAAACTACT	GTATTCAAAG	GCCGGACTAA	720
TTyCAGCGTA	GGAAGCCCAA	CCAGTTAAGA	aGCCCCAAAA	TTTCGTTGCC	AATGACCATT	780
GACTATTTAC	AAATAAAAAG	GAGCTTAAAA	TTACGCCTAT	CGCTGCCCCAT	TTGACTTTGC	840
GATGATCTAA	AAAGGTTTGT	AGCTTTTGAA	TATCCATTAA	GTAGCCACAA	TAAAAGAAAG	900
GGAAGAAGCC	CAGAGTCCGT	CCTAAAGACA	AGAAACCTTC	CACAGGTAAA	TAGCCTGAGG	960
CGCAgCcGAT	AACAATGATG	ATTGGCAACA	TGATTTTCGG	CTGAACGCGC	TGAAGAAAGA	1020
TGGTAATTAA	GTAAGAACTT	GCTAAGGAAA	GTAAGTACCA	AGGAACGCCT	GTTTCGTTCA	1080
GTAAATCAAA	ATTCATTTTA	TTTGCAAAGG	CGAAATTAGC	AAGCGTAATC	AACAACCTCG	1140
CTAAGAGATA	AAGCCAAACA	AAGGAAAGTA	AGCGTAGATA	TTTTCGTTGT	TTAATGGCAT	1200
GTTTACTAAA	GAGTCCAGAG	ACAAAAATAA	AGGCTGGCAT	ATGGAATGAA	TAGATCCAAA	1260
AGTTGATCGT	TTCCATGAGT	CGTGAAGGAG	AAGCTTCACG	ATCTAAAATA	TGCCCCAAAA	1320
TCACTAAGAA	AATTAAAAAA	AGTTTCACGT	TGTCCCATTT	ATAAATTCGC	TTGTCGTTTT	1380
CCATAACGAT	CACTCCTTTT	AGTAGAAAAC	AGTTCACTTT	TGTTTAAGGC	AGAAGAGGAC	1440
AATTTTTCCA	TTCATTGAGT	AGATAATTGT	AGATACGAAC	CGGCAGCCGT	TAAATGAGCC	1500
CCCACATACT	TGCGCTTAAA	ACAAAAAATT	AGGCAGGATG	TTAGTCTTAA	ATTATACTGA	1560
AAGAGTGACG	ATCCTCCAAA	GATTTTGT	ATTAAGAAGT	GAAAGATGCG	GTGAAAATAA	1620
ATGATAAAAT	TTCAAAATGT	TTCAAAAGTA	TATCAAGAAA	ATGCCGTTTT	GTCCGATGTG	1680
AATTTAACAA	TCCAACAAGG	CGAATTTTTT	GTATTAGTAG	GGCCGAGTGG	TAGTGGAATA	1740
ACGACGACTT	TGAAAATGAT	TAATCGGCTC	ATTGAGCCAA	CGACAGGTGA	AGTGGCTCTC	1800
AATGAACAAG	CGGTAACCAA	TTATCCTTTA	CGTGAATTAC	GAATAAAAT	TGGCTATGTT	1860
CTTCAGCAAA	TTGCCTTATT	TCCTAACTTA	ACCGTGGCGG	AAAATATTGA	GTTGATTCTT	1920
GAAATGAAAA	AATGGCCCAA	AGAGCAACGC	AGAGAGCGAA	CCATTGAACT	ATTGAAAAAA	1980
GTTCAGCTGG	ATCCTGAGGA	ATATCTGCAT	CGTAAGCCAG	CAGCGCTTTC	AGGAGGCGAG	2040
CAACAACGTA	TCGGGATTTT	ACGCGCCATT	GCAGCCGAAC	CAGAAGTGAT	TTTAATGGAT	2100
GAACCTTTTA	GCGCGCTAGA	TCCGATTTCA	AGACATCATC	TTCAACGTTT	GGTGAAAGAA	2160
CTTCATCAGG	AATTAGCGAG	TACCATTGTT	TTTGTGACCC	ATGATATGAA	TGAAGCTTTA	2220
CTGTTAGGAG	AGCGCATCTG	TATTATGGAC	ACTGGCAAAA	TTATTCAAGT	CGATACGCCA	2280
GAAGCCATTC	AAAAACATCC	GAAAAATAAT	TTTGTGCGTC	AATTTTTCAA	ACAATCTTCA	2340
CCAGAACTAG	CGAACTATTG	TGGCGCTGAC	TTGACTGCAT	TTTLAGAAGC	CGCAGATGAG	2400

ACGATTGAGC CAGTGCCGAC TGTTTTCATG GAAACCAATT TAAAAGAGAT TGTCCAAC TC	2460
ATTAATGCGC AAGAGAAAGC AAATTTAGTA GAAAAACAAC AAGTTT TAGG GTGTATCACC	2520
TCAAAAACGA TTA CTGTTT TATGGAACAG CTGTTGGAAA GCGAGGGACA GTAGATGCAA	2580
GCACTGATTG AAAC TTTTTT CGCTCGAAAA GAAGAATATC TGCAAGCGTT AATCGAGCAT	2640
ATGCAATTGT CCTTATCTC GTTATTATTC GCTGCGTTGA TTGCAATTCC TTTAGCCATT	2700
TATTTAACAA GTCATCGTAA GTTAGCGGAA AGTCTGTAC AAGTGACTGG GATTTTCCAA	2760
ACCATCCCTT CTTTAGCCTT GCTAGGCTTA CTGATTCCGT TAATTGGTAT TGGTCGACCA	2820
CCGGCAATTG TTGCGCTAGT GATTTATGCC TTATTCCCAA TTTTACAAAA CACCTATACT	2880
GGTTTGACGG AATTGATCC TTCGTTAGAA GAAGCTGCAG AAGCATTGG GATGAGTAAA	2940
AAAGAAAAAC TGTTTAAGTT TGAGTTGCAA ATGGCGTTAC CATTATTAT TTCTGGGATT	3000
CGAACGGCTA CCGTTTTAAT TATTGGTACG GCCACGTTGG CGGCGTTAAT TGGAGCAGGT	3060
GGTTTAGGGA CCTTCATTTT ATTAGGAATT GATCGAAATA ATTTATCGTT GATTTTTATT	3120
GGAGCCCTTT CATCGGCTGC ACTAGCAGTC TTATTTAATT ATGGAATTCA TTGGTTGGAA	3180
AAAGCGAATG GCGGCGGACT AATCATTGGC GGTACTATTT TAGGCCTTTT GTTAGGTGGT	3240
TCATTTTTCT GGAATCAACA AACTTCTTCA AAGGAAAAGC AGCTGACAAT TGCTGGCAAA	3300
TTAGGAGCGG AACCAGATAT TATTATTAAT ATGTATAAAG CCTTAATTGA AGAAAATAGC	3360
GACATTCAGG TGACGTTAAA GCCAACTTT GGCAAGACTA CTTTTTTATA TAATGCCTTG	3420
AAATCAGATG AAATCGATCT TTATCCGGAA TTTACAGGAA CAGTGATCGA AACCTTTTTA	3480
AAAAATCCTC CTCAGCTAGA CAATCAACCT CAAGTAGTCT ATGAAGCGGC ACGCGCTGGC	3540
TTGAAAAAAC AAGAAAACCT TACTTTGTTG AAACCGATGC GCTATGAAAA TACGTATGCC	3600
GTCGCAGTTA AGCGCTCATT TGCAAAAGCG CATCAATTAA AAACGATTAG TGATTTACAG	3660
AAAATTAGCA ATCAGTTGAA AGCAGGTTTT ACGTTAGAAT TTATTGATCG GCAAGATGGT	3720
TATAAGGGAT TACAAGAAAA ATATCATTTA AATCTGAATG TCCAATCGAT GGAACCGGCG	3780
TTACGCTATC AGGCGATTAA TAATGGTGAG GTAAATGTGA TTGATGCCTA TTCAACGGAT	3840
AGTGAATTGA GACAATACGA TTTAGTAACG TTAGAAGATG ATCAAGCGCT ATTTCCGCCT	3900
TATCAAGGAG CACCACTTAT TAAAACAGCA ACGTTGGAAG AATATCCAGA ATTAGCAGAA	3960
ATATTGAATA AATTGGCAGG AAAAATTAGC GAAGAAGAAA TGAGTGAAAT GAATTATCAA	4020
GTCAATGTGG AAGGGCAAGA TCCTAGCATA GTGGCGAAAG ACTATTTAAA AGAAAAGAAT	4080
CTTCTTAAAT AAGTATTTTT TACTGAAAA GACTGGTATA TAAACATTTA TCCAGTCTTT	4140
TTAGATAAAA TAAAGAGTTT AATGTGAAAA GTACCTGTAT TTTTACTCAT TTTTGGATAT	4200
AATAAAAGAA ATGAATGGAG AGGAGTGGTC TATATGAAAA AAGCAGAACG TCAACGTTTG	4260
ATTAAACAAT TAATTATGCA ACAAGAAATT GAAACACAAG ATGAATTAAT CACTCGTTTG	4320
GAAGAAATTG GCGTCCGAGC CACTCAAGCG ACTGTTTCAA GAGATATCCG TGAAATGAGT	4380

ATTGTTAAAA	CACATGGTGC	AGATGGTCGT	GTCAAATATG	CAATTTTTTC	TCAAGCACAA	4440
GGCACAAGCA	GTGAAGAGAA	ATTACGAGAA	TCAGTAAAAG	ATTCAGTTGT	TCGAATGGAG	4500
CGAGTACAAT	TTATTGTCAT	CTTGCATACA	GAGATGGGAA	ATGCCGATGT	GGTGAGTAAC	4560
TTTTTAGATG	AAGTAGCTTA	TCCAGAAGTG	GCTGGTACAG	TGGCTGGTGC	GGATACAATT	4620
ATTGTGATTA	CACGTTCAAG	AGAGGATGCG	GAACACTTTA	TCGAACGCAT	CGAAAATATG	4680
ATTTTTTAAA	ATTGAAAGCA	TCAATAGAAG	GGAGTTTTTT	TCTTGGCAGA	ATTGAAAGCG	4740
CTACAAAATG	GTTCCGATAT	TCGGGGAATT	GCTTTGGATA	CGGAAGAGCA	AACTGCAACA	4800
TTAACCGBAA	CAGCAGTAGC	CGAGATTGCG	GTGGGCGTGG	TTCGTTGGTT	ACAGGATAAA	4860
AAACAGCTGC	CACGAAAGGC	ACAACAGCGC	CTGACAATTG	CAATCGGACA	TGATAGTCGT	4920
CTAACGGCGG	AGcatTAAGC	AAGCATTAGT	AGATACGTTT	CTTTCATTAG	GTATTCAGGT	4980
GATTGACGTT	GGTTTAGCTA	CGACGCCTGC	TATGTTTTATG	GCTACTCAAT	TTCCAACGCT	5040
TCAATGTGAT	GCAGCGATTA	TGATTACAGC	TAGCCATTTG	CCTTATTACT	TTAATGGCTT	5100
GAAGTTCTTC	ACAGCAGAAG	GTGGAGCAGA	GAAAGAAGAT	ATTCGGTATA	TTTTATCTCA	5160
TACTGATCCG	CTGACGGCCA	ATGAAAATGG	CACACTGATG	AAGCAAGAGT	TACTGCCAAT	5220
TTATGCGGAG	CATTTGGTGG	AAAAAATCCG	CCAAGGAATT	CATTCGCCAG	AAGAGAAACC	5280
TTTACAGGGC	TTTCGAATTA	TCGTTGATGC	GGGTAATGGG	GCTGGGGGCT	TCTTTGCGGA	5340
GCAGGTTCTT	CAAGTTTTGG	GCGCTGATAC	GACGGGATCA	CAGTTTTTAG	AACCAGATGG	5400
GCATTTTCCT	AATCATGAC	CTAATCCAGA	TAATTCTGAA	GCAATGAAAA	GTATTCAAAC	5460
GGCTGTTTTA	GCTAATCAAG	CAGATTTGGG	CATTATCTTT	GATACAGATG	TAGATCGTTC	5520
GGCCGTGGTG	GACCAATCAG	GTGAGGTACT	AAACCGAAAT	AATTTAATTG	CTGTTTTAGC	5580
GGCAATTGTT	TTAAAAGAAG	CGCCTGGTAG	TTATATTGTG	ACAAATTCGC	CAACCTCAAG	5640
TCATTTAAAA	ACATTTATTG	AAGAAAAGG	CGGCCAACAG	ATtCGTTATA	TTTCTGGTTA	5700
CCGAAaTGTC	ATTAATAAaA	TGATTGAATT	GAaTCATGGG	GGCTTTCAGA	CACCGTTAGC	5760
CATTGAAACC	AGTGGTCATG	CAGCGTTTCA	AGAAACTAC	AATCTGGATG	ATGGTGCATA	5820
TGTCGTGGCG	AAAATTTTAA	TGTTGTTACC	AGAATTGAAA	CAAAACAATC	AAACGTTAGG	5880
AGATCTGATT	GCTACATTGA	AGCAACCGGC	AGAAACCAAC	GAATTTGCT	TCAAAATTAC	5940
TGCGGAGGAT	GTTACTTGTT	ATGGGCAACA	AGTCTTGCGA	GACTTTGAAT	TGTTTGTAGA	6000
AAATCAAGCA	GACTTTGCTG	TGGACCGTGA	AAACCAAGAA	GGTGTTGCGG	GAAACGTGTC	6060
TGGTCAATAT	GGCAGTGGTT	GGTTTTTGTA	CGCTTAAGTC	TGCATGAACC	GTTGCTTGTT	6120
CTTCAGGTAG	AGAATGACCA	AAGCGATAAA	AATGCGTGTG	TCATCGAAAA	AATTGCCACT	6180
TTTTTACAAA	AATACGAGGA	AATAGATAGT	CAGCAAATAG	AAAAATAATC	TAGGAAATTA	6240
AAGGTGTTCT	TATGGAAAAT	GTTTTAAAAG	GAAAACACGT	TGGCTTTTCT	CATTTTAGAG	6300
AAGAATATAT	CGAACATATT	GCAAAACAGC	AATGGGATAA	TGAATTGCTT	CGGCATTGTA	6360

GCTGGGATGC TTTACATCCG TGGGGGCTTG AAGAATGGAA AGATTTCACA ATTAATAAAG	6420
GAGAAGATGA TCGATTTCTA TTTGCTATTT TAGAAAATAC GACAGAGGAA TTTATTGGTT	6480
GGGTTTCGTT GTCTGATGTA CAGTTGAAAA ATAGAGGGGC CAATTTAGGG ATTGCTATTT	6540
TACAAAAAGA GCAACGAGGT CAAGGATGTG GCTTTGAGGC GGTGTCTTTA ATTTGTAAAT	6600
TTGCTTTTTA TGAATTGGGT TTGCACAAAA TCAGATTAGC TGTCAATAGT AACAATCAAA	6660
AAGCGATTCA TGTTTATGAA AAAGTAGGCT TCAAAAAAGA AGGAATTGAT AGAGAAGCGT	6720
TATTTCAAGA TGGCCAATGG CTGGATATCT ACAATTACGG AATTTTACAA AAAGAGTGCG	6780
TACAAATGAT TAAGGCTGAG AGTTAGCTCA GGAAGTGTG AACACCAAGA CACTGAGCAA	6840
ACGCAGATAT GCCGTTCTCA GCAGTGAATG TCGGGTAAAA TAAAAAGCA GAAGAAATCG	6900
GCGGTTACCG TCCGATTTCT TCTGCTTTTT TTTACCATAG GCCAGGTGTT GCAGTCACAA	6960
CGATTTGTTT TTTTGTAAC GGATGGAGTA AATGAAGTTG TGAAGCATGG AGCATTAAAC	7020
GAGCTGCCTG CGATTTTGGG TTGTATAACG GATCACCAAT AATTGGATGC CCTTCGCTAG	7080
CTAAATGGAC CCGAATTGGG TCGGTCCGAC CTGTTTCTAA AACACAGGTA ACTGCGGTTT	7140
GATTATTAGA AAGAGCTTGG ACCACTTGGA CATGTGTAAT CGCTGAATCC CCTTTTTGAG	7200
GATCAATCCG TCGTTTCCGG CGATCATGTC GATCACGACC GATCGGCTGA TTGATTGTCT	7260
GCTCTTTTTG GCTAAAGTGA CCAGCTACAA TCGCTTGATA TTGGCGGTAA ATTTGTTTTG	7320
TTTCTAATAA ACGACCTAAA ATAGGTAAGA CAAAAGGATT TTTAGCAAAG ACAATCGCAC	7380
CACTGGTTTC TTTATCTAAT CGATGAACGA CATAAGGGAC CTGATTTTTC TCGGCTAAAT	7440
AGGCAGCAAC ATGATTTAAC AAGGTGTCTG TTTCTGTGG CTGATTTGGA TGGGTTTAA	7500
TGCCATAAGG TTTATTTACA ATCATTAAT GTTCGTCTTC ATATAAAACC TTCACATGCG	7560
TTGCTTCCCC CAAATGAATA GTGGGACGTA CATAATCGGT TTCTTCAAAC GTTAAGGTGA	7620
CACAATCGCC AGCTTGGAAT GGAATGAA AAGCAGCTGT TTCTCCGTTG ATTTTGACAT	7680
TTTT	7684

## (2) INFORMATION FOR SEQ ID NO: 17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32768 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

CGGAACACCT GnCGCCTATA AAAAGGTTGG CCAACAAACC CTGATGGCTG GGGCGATTGA	60
TAAACCGAAC ACTTACACGG TCAAAwTTga TGTAAGAAcC GAGGGTTCTT TAAAAATCAA	120
AAAAGTTGAT AAAGAATCAG GTGCTATTGT ACCAGGAACG GTTTTCCATT TAGACTTTGG	180
AAAAACTTTA CCTGCAAAAG ACGTGACTAC TGACAAAGAA GGCATTGCTA CATTGGATGG	240



GATTCTCAT	GGTACAAAGG	TAACCATTAC	CGAAAAATCG	GTGCCGGCGC	CTTATACGAT	300
TGATACCATA	CCCATGACTG	CTACGATTAA	AGCAGGCGAA	ACCATTTTTA	TGACTTCAAA	360
AAATACGCGG	GAAAAAGGTC	AAATCATTCT	GGATAAAACC	GGGATTGAAA	CAGGAAGTGA	420
TCTTTGGAAC	GACAATTACT	CTTTAGCTGG	AAACACGTTT	GCTATTCGCA	AGGACAGTCC	480
TACTGGTGAA	ATTGTTCAAG	AAATGACCAC	GGATGAAAAA	GGTCATGCAG	AAACACCAAA	540
AGAGATTGCT	AATGCATTGG	AATTGGGAAC	TTACTACGTG	ACCGAAACCA	AAGCATCTAA	600
TGGCTTCGTG	AATACCTTTA	AAGCAGTAAA	AGTCGAATTG	AAGTATGCCA	ATCAAACCGT	660
GGCCCTTG TG	ACCAGCAATA	TCAAAGGGCA	AAACCAAGAA	ATTACTGGGG	AAACAACTCT	720
GATAAAAGAA	GACAAGGATA	CTGGAGATAA	AACACAAGGC	AGGGCAGTCT	TTGAGGGGAC	780
TGAATATACC	TTGTTTACTG	CTAAGGATGG	AAAAGCGGTT	AAATGGAGTG	AAGCTTTTAA	840
ACCAGAATTG	GTAAAGGGAA	CGAAAGCTTC	TGATGAAACA	GTGACTCTGG	CTTTAGATGA	900
AAAGAA TCAA	GCTGCCGTTa	AACAcTAGC	CATCaACGAG	TACTACTGGC	AAGAAACCAA	960
AGCACCTGAA	GGATATACCT	TGGaTGaAAC	GAAGTATCCT	GTwTCyATCA	AAAArGTwGA	1020
TgATAACGAA	AAAAaTGCCG	TrATTACyCG	AGATGTTACG	GCAAAAGAAC	aGaTTATTCCG	1080
TTTTGGGTTT	GATTTCTTCA	AATTGCGyGG	ATCrGCyGmT	GGCACTGCCG	AAACTGGGTT	1140
TAATGACTTA	ACCTTTAAAG	TGTCGCCATT	rGAAGGGACC	AATGAGATCA	CAGGTGCTGA	1200
AGATrAAGCA	ATCACAGCTT	ATAATGAGCA	ATTAGGCTTT	GATGGCTACG	GAAAGTTTGA	1260
AAATCTTCCT	TATGGGGATT	ATTTACTTGA	AGAAGTAGAA	GCGCCAGAAG	GATTTCAAAA	1320
AATTACGCCA	TTAGAAATCC	GTTCTACTTT	TAAGGAAAAC	AAAGAAGACT	TTGTGAAGAG	1380
TGAGTATGTC	TTTACCATTA	CCGAACAAGG	CCAAAAGCAA	CCAATTAAAA	CGGTGACCGT	1440
TCCTTACGAA	AAGCTGACGA	ACAAAGCATT	TTCTGTTAGT	TTGAACCGTT	TGATGCTCTA	1500
TGATTTGCCT	GAGGAAGAAG	ATAGTTTGAC	TTCTCTTGCA	ACTTGGAAGG	ACGGAAATAA	1560
AAAATTGAAT	ACCCTTGATT	TTACCGAGCT	AGTTGATAAA	TTGAGATATA	ACTTGCAATGA	1620
AATCAAAGAA	GACTGGTATG	TCGTAGCTCA	AGCCATTGAT	GTGGAAGCCA	CAAAAGCTGC	1680
CCAAGAAAAA	GACGAAAAAG	CCAAACCAGT	GGTAATTGCC	GAAACAACGA	CCACTTTGGC	1740
GAACAAAGAG	AAAACAGGTA	CTTGGAAAAT	TCGGCATAAA	TTAACGACCG	AACAAGTTTT	1800
GGATAAAACT	ATCGTCTTGT	TTAACTATGT	GTATGAAAAC	AAGGAAGCTT	TTGAAGCAGG	1860
CGATGAGCCA	GTAGCGAAGG	ATGCTAGCTT	GAATAATCAA	GCCCAAAC TG	TTAATTGTAC	1920
GGTGAACG T	CATGTTTCCA	TCCAAACAAA	AGCCCACCTA	GAAGATGGTT	CTCAGACCTT	1980
TACTCATGGT	GATGTGGTGG	ATATGTTTGA	TGATGTATCG	ATTACCCATG	ATGTGCTGGA	2040
TGGATCAAAA	GAGGCTTTTCG	AAACAATTCT	GTATGCACTA	CTACCGGATG	GTACAAACAA	2100
AGAAATTTGG	AAATCAGGCA	AGATTGATTA	TGAAGTAAAT	GATAAAGAAT	TTACCAAAAC	2160
TATGCTTGCG	GAAAAAGTAG	ATACCGGAAA	GTTTCCAGAA	AGTACTAAGT	TTACTTTTGC	2220

AGAAATCAAT TATGATAAAG ATGAGAATAT CAATGGGAAA CACAATGAGG ATCTAAAAGA	2280
AAAAAATCAA ACCTTAACGC CCAAAAAAGT TCCAACCACA CCACGTACGC CGAAACAACC	2340
GGAAACACCA ACTGTTCCAA GTGGCTATCA AGAATCTAGT CCCACAGTGA AGACATTCCC	2400
TCAAACCTGGG GAGAAAAATT CCAATGTTTT ACTGTTTCATT GGTTTTACTT TGATTTTTGC	2460
GACGGCGGGC TATTATTTCT GGAATCGCCG TAACTAAGGT GATGAGATGA TACAAGTAAA	2520
TCCTAAGAAC CAACGTGCGC GACCGCCTCC TAACAAAGTG GTCGCGCTAT TTTTGTATCC	2580
AAACTAACA ATGAAAAAG GAAAAGAAGA TGGAATTGAA ATTTGTTGTG CCAGATATGG	2640
CCGAAACGTT TGGAAAGTTA AGTTATGCCG GTGAAGGAGA AATTTTAAAC GAAGGATATG	2700
GACGGAATAC TACGGTGATT GGTGCGAGTT ACCATCTATA CTCCAGCAAA CAACGGGCCC	2760
ATGATATTGA GGTGGTAGTA GCTGCGGAAG CCGGTGAAAA GGATTTTGAT CAAGACCAAC	2820
CGCTAAAAGC CGTGAATCCC CATTTGGTTG CCAAAGGCTA TGAGATTGAA AATCGTGGGT	2880
TTACTGATTA CGTGTGTAT GTTGATGATT TAGTGAAAGA ATAGGAGGAA AaGAAGATGA	2940
GATTAACAGA AGGTATTGTC GTAGATTCAG TATTAACATT TGGAAAGTTG CGTTTTTCCG	3000
CATTACGCCG AGAAGTACGG AAGCAAAATG AGGATGGAAC GGTTAGTAAT GAGGTAAAAG	3060
AACGAACATA TAATTTGAAA TCCTCTGCAC AAGGGCGAAT GATTCAAgTT AGTwTCCCGG	3120
CTAATGTGCC CTTACGCGAG TTTGCTTATG ATGCAGAGGT AGAGTTGGTC AATCCAATCG	3180
TGGATACAGT CGCCAATTAT GTTTTTCGAG AAGGAAC TAC TGTCAATTGG TTTATTAAGG	3240
CAGACGATTT AGTcTTGAAA CGACAACCAA ATCAAGGAAA TTCTACGAAC CAAATTGAAG	3300
GAAAGAAATA GGTGAAGATT ATGGAAATTA ATGTTGAACA TATTTTAGAT TGTCTCGATC	3360
AATATGGAAG AGGTGAGTTA ACTGAGGAAC AACTCGTTAA AGCATTAAAC TATGATGAAA	3420
AAATGTTTCT GATTATGCAT CAAGGTTTAC TTGGAGAAAA TAGCGACAGT GAAGACGACT	3480
TTGATGTACA AGACTGGTTA GAAGAAGAAG AGTCCTTCTT TATGGTTGTC GAAGTAAATG	3540
AAACTTATG TCGGCAGGCA GAATCCGTCT TGGAAGAAAT CGGTGTGGAG ATGCCAGATG	3600
CCATTGAAGG CTTTTTAAAT CAATTGGTTG AAAAGAAACA ATTCCCCGTA GTAATAAATA	3660
GCTAGCATAA CTATATGAAA ATGCCCAATG AAATCCTATC GGGCGTTTTT TGATCGGAGG	3720
TGGAAACCAT CAAAATGTAC AAAGAACATC GCATACGTGC AAGAGATCAG CATTTGGTTT	3780
ATCACTTTAT TCTTGGCTGG CTAATTGCTT TGTTAATCAG TTGGATGGGT GTCTTTTATT	3840
TTCAAGAGTT TAGACAATTT GATATTTCTA GGGTATCACT TTCTACTATT GAAACTGTTT	3900
GGTCCATGAA GGAGTTAATC TGCTTACTTG GAAGTCTCGG TTTTTCAGGA GCTATGCTGT	3960
TGCTTTATAT TCACTTTTTT CCTGATCATT GGCGAAGCTT GTGGCATCGG CAAAACTAG	4020
CTCGGATGAT TCTAGAAAAC CATTGGTATG AAGTGAAACA AACTCAAAGT GAGGGCTTTT	4080
TCAAAGATTT AAATAGTAGT CGAACCAGAG AGACTATCAG TTAATTCCCC AAAATCTATT	4140
ATCGAATGAA GGAGGGTTTA CTTTCTATTC GCGTTCAAAT TTCCTAGGG AAATATCAGG	4200

AGCAGCTATT GAAGTTGGAA AAGAAGCTAG AAAGTGGATT GTATTGTGAG TTAGTAGAAA	4260
AGGAACTCAA AGACTCCTAT GTGGAGTATA CTTTACTGTA TGATATGATC GCCAATCGAA	4320
TTGGGATAGA CGAAGTAGTG GCAGAAAATG GGAAGTCTGCG ATTGATGAAA AATCAAGTAT	4380
GGGCTTATGA TTCTTTACCC CATATGTTAA TCGCTGGTGG CACAGGTGGG GGAAGACCT	4440
ATTTCTCTCT CACCATTATC GAAGCACTAT TGAAGTCAGA TGCAGAATTG TTTATCCTCG	4500
ATCCCAAAAA TGCAGATTTA GCTGATTTAG GTACGGTGAT GCCTCACGTA TATTCTCAAA	4560
AGGAAGAAAT TTCTGCTTGT GTGGAAGATT TTTATGAACG CATGATAGCT CGTAGTAAGG	4620
CAATGAAAGA AATGCCAAC TACAAACCAG GAGAAAATTA TGCCTATCTT GGACTTCCAC	4680
CAAACTTTTT AATCTTTGAT GAATACGTGG CTTATATGGG TGCGAATCGT TTTCTACAA	4740
GCATCGAATA GATGTGAAAT GGAAGCCTAT TGAGATTATC AATAGTGACA ACTAATGTTT	4800
TGAGAGATGG AATGAAAGCC TTCAAGTTGA GCTGAAACAT CTCCGCTAAT ACTCCTACAT	4860
GCTCTAATCA ATGAAGTAGA TTGGTTTTGG TATGAAGCTA GGTGAAGTCG TCTGAAATCT	4920
GCCTAAGTGT TTGAGCATAA GGCTGACTAT AGGGCGGGCG GCTGAAAAAG ATAAGTATGG	4980
TAAGAATGTC TGAAGATGAC TGACGAATCT GCGAATGTAC GGGTCTAAAA GGCTGATCTT	5040
GTGAAAGAT AGGAACATAC GGACGTATGG TGAATACCG ACGAGtaagG TTCCATGATA	5100
TGAAATCGG GATAGCTAAC GAAGAAGCTA TAATTTCACA GGCTCATAGT AGATACCTAA	5160
GTTATACAT GATAGCTAAA AACATTGGAA CGATGTAAGC AGAAACTAT CAAATAAATG	5220
ACTACTTCTA GATAGGTGCA TATAAGTAAT AACGAAATTG CTTTGTCTCT GTGAAAGTAA	5280
GGGCACAGTA CCGATGAAGC GTCTAACAAA CGTGGAGGGA TAGCCCTAAG TCTATTTAAT	5340
AACAATGAAT GTTGAAACAG ATGAAGTTCA AGGATCTGGT AAGAACGTGG GGAAATCCGT	5400
TATGGAGGTG CCACGCTTGA ACACAAAAAT ACGTTACTGG GAATATTACG ATCTACAAGA	5460
AACATTGAT CAGCTATTTA CTCAAAGTAA AAACGGGAAA AAGTTTTATC AATTATACGA	5520
GTTAATCATT TCTGAAAATA ATATTCTTTT AGCATACCGC ACTATCAAGG CAAACAAGGG	5580
GTCTTCGACA CCAGGAACGG ACTCCTTCAC AATTGATAAC TATAAGGAAA TGAaCCAAGC	5640
AGAGTTTATA CATCTTATTC TCAGTCAGTT AGAAAATTAC AAACCAAAAT CCATAAAGCG	5700
AGTAATGATA CCTAAACCCA ACGGCGAGAA ACGCCCATTA GGAATTCCaT GCATGATTGA	5760
TCGAaTTATC CAACAGATGT TCAAGCaGGT ACTGGAGCCA ATATGTGAAG CAAAGTTTTA	5820
TGAACATAGT TACGGGTTCA GACCATTGAG AAGTGCAAAA CATGCTCTAG GTCGCATTAT	5880
GTATCTCATA AACATCAGTA AGATGCACTA TGCAGTAGAC ATTGATATCA AGGGTTTCTT	5940
TGATAATGTA AATCACCGTT TACTGATAAA ACAATTATGG AATATAGGAA TTTGTGACAA	6000
ACGAGTCTTA GCTATTCTTT CAAAATCTCT GAAATCACCC ATTCAGGGAG AAGGTATTTT	6060
GAGCAAAGGA ACGATCCAAG GAGGTATCAT TTCTCCACTA TTATCTAATG TTGTTCTAAA	6120
CGACTTAGAC CATTGGGTCT CAAAGCAGTG GCATACCTTT GAAACAAAGT ATCCCTATAC	6180

GAAAGGTTAC AATAAATTTT	GTGCTTTAAG AGATACAAAT	CTAAAGCAAG GCTATATTGT	6240
CAGGTACGCA GATGATTTTA	AAATCATGAC TAATGATTAT	CCGTCCGCAT TAAAATGGTT	6300
CCACGCTGTC AAGCTCTACC	TAAAAGACAG GCTTAAATTA	GATATCTCCA ATGAGAAGTC	6360
TAAAATCGTC AACTTGCGTA	AACGCAAATC GGAATTTTTA	GGATTCACCA TTTGCGTGAA	6420
ACAAAAGGGG AAAAATGGGT	TTGTAATTCG CATATTTCTG	ACAAAAGAA AAATCAAATT	6480
AAAAAAGAGA TTAAACAAAG	AATTAAAGAT ATTCAAAAGA	GCCCTACAGC ACAAATGCT	6540
CTCCTATTTA ATGCTCTAGT	GCTTGGTGTT CATAATTATT	TCAGGTATGC CACTCATATT	6600
AACTTAGATT TGAATAGGAT	GAAGTACGTT CTTTCATTCA	CTACGTATAA CCGATTATCA	6660
AATTGTAGTA CAAGATGTTA	TCCCAACAAA GCTTCCCCTA	CTTACAAAAA GTTTTACTCG	6720
ACAACGATGA AAACGTATAA	AGTAGCCGGA GTCTATTTGT	ACCCTCTTTG CGACGTATCT	6780
ACaAAAACAA TTTACGGATT	TAATACGGAA GATACACCGT	TTACCCCGTT TGGtCGTCaA	6840
aGGTTGGAGC ACAAAGCTT	ACAGTCTCTT GTATACCAAG	AATTAAGAAA ATTGATGGAA	6900
TCGAAAATAC TGAATAGAAT	GGTTGAGTAT TTAGATAATA	GAATATCTCG TTACTCAATG	6960
AAATCAGGCA AGTGTGAGAT	TACAAAACAG TTTTGGCCAG	CAAAAGCCGT TCACTGTCAT	7020
CACTATTTAC CAAAAAGTCT	TGGTGGAGAC GACAAATTTG	ATAACTTGAG AATTATTCAT	7080
AAGGATATCC ATCTTTTGAT	CCATACCACA AATAAAATGA	TAATCGATCA CTACGTCAAT	7140
GAAGTGAAAC TGCTGCCAGA	GCAATAGCA AAAATCAACC	TTTATCGAAA AATGTGTAAT	7200
CTGCAAAACA TTCAATAGAA	AATAAATAGA TGGAACGCCG	TATGAGGTGA AAGGCTCACG	7260
TACGGTGTGG AGCGGGGGAA	AAGATGGAGA TTATATCAAA	TTCTTACCTA TCGTTATAAA	7320
TGTTAACGAC GAAAGAAAGT	GCGGTGATTT TAAATAAGCT	CAAACAAATC GTGATGTTGG	7380
GTCGTCAGTC TGGTTTCTTT	CTAATTCTGG CTTGTCAAAG	ACCGGATGCA AAATATTTAG	7440
GAGACGGGAT TCGCGATCAA	TTTAACTTCC GAGTGGCTTT	GGGTCGTATG AGTGAAGTCG	7500
GGTATTCCAT GATGTTTGGA	GAAGTTGATA AAAATTTCTT	TATGAAACGC ATCAAAGGTC	7560
GAGGCTATGT GGATACTGGA	GGGAGTGTGA TTAGTGAGTT	TTATACGCCA CTTGTGCCAA	7620
AAGGTATGA TTTCTTAGAA	TCCATCAAAC AAGTAGCACA	ATCAAAAGAA AAATGACAGA	7680
TACAGAAAAA TTTAATCTAC	TAGTTCATCG GGTAAAAGT	ATGATAAACT AGCTTTGAAG	7740
GTGTCAGGGA GGTGCAATAA	AAATGACAGA GGATGATTCT	TTACAATATG CCaAGGAGAA	7800
CAAGAAGGAA ATTATTTCTT	CTGTTATTGA TGGAAAAGAA	AAAGAAGAGG AAAAACTGC	7860
CATTTTATG GCTGGAAGTC	CTGGGGCTGG CAAGACTGAG	GCTGCCCAA CTTTAACGGT	7920
ATTAAATAGT AATCTTTGTG	TGATTGATGC TGATAAGTTT	CGTGTATTAT TTCCAGGTAA	7980
TGTTGGGAAT AATTCAGATG	AATTTCAACG AGGATCATCA	CTCTTGGTAG ATGCAGCTCT	8040
CGATTTAGTT CTTAAAAAAG	GCTATAGTTT TATTCTTGAT	GCTACTTTTG CGACCTCTAA	8100
GGTAAACAA AATATTGAAC	GGGCACTGAA AAAGAATTAC	AATGTGCTTG TCTATTATGT	8160

TTATCAAGAT CCATTTATTG CTTGGGATTT TACCAAAAAA TGAGAAGTGG TTGAAGGGCG	8220
TTTTGTTTCT AAAGAGCGTT TTATCAATGC TTTTTCCTCA TCACGAAAAA ATTTGATGCG	8280
TGTGAAAGAG CAAATTTCAA GATCAAGTGA CAATCAATAT CTTAGTGAAG GACTTTCAA	8340
ATACGATTTC TGATTTTCTG ATGGATATTG ATAATGTTGA GTTGACATTG CCAATCAGTT	8400
ATACAAAAGA AATATTGGAG GAAGAGCTAC ATGACTGAGA AAGAAAAAGT TGAGGAAATT	8460
ATGGAAAAGT ATAACCGTAA CTTTTCACACA TTACAAAAGA ATGCTTCTGC GAAAGAGCTG	8520
AAGACAGTCT TTAAATTTGT AGCAGATGAA TCCAACCGAA AACACCGTGA ATTGATTGGT	8580
TTGGATAAAG AAAAATAAAT ATGAAATGAG AATAGTTGTA GAAGCCAAGA GCAAAAATGT	8640
TCTTGGTTTT TTTCTATGCA ATGAAAGTAG ATATGGAAGG AAAGGAAATC ACAATCAGCA	8700
ACTCAACTGA TAGGCACTGA CCGCGGAGGA TTTGCGAAgc AGAAATCGTC CGTGGTCCTG	8760
TGCCTGTTCA TGGACGGCGC TAGcgGCCAT GATGTCTCAC CCCCCGCATC TAACAGGGGG	8820
GTACAAAAC ACAAGAAATA AGCAAGCAAC GATTAGAAT TGCTGGTGTA GCAAGGATAT	8880
TAAGTAATTT GCATAGGTGT TAACTCATTT GAAAAAGTTA ACACCTGAA TTTGGAGGTG	8940
AGAACTTGGT ACAACGGAAT TTAGATTATC GGTTATTAAA AGATCGTCGC AATGAATATG	9000
GTATTTGCGA AAATAAATTA GCTACCGCTT GTGGGCTTAG TCGACCCTAT TTAAATCAGA	9060
TAGAAAACGG TGGCGTAACT GCCTCTACAA AAACCATGAG AAAAATTTTT AACCAACTGG	9120
AAAGTTTTAA TCCCGATTTA CCTTTATCGT TACTGTTTGA CTATGTAAGG ATTCGTTTCC	9180
CCACAACGGA TGCACGGAAT ATCATTCAAG AGATTCTCCA CCTGAAATTT GATTATATGC	9240
TCCACGAAGA TTACGCCTTT TACTCCTATC AAGAACAATA TGTCATGGGC GATATTGTGG	9300
TGATGTTGTC TCATGAAGAA GATAAAGGAG TTCTTTTGA ATTAAGGGT CGTGGTTGTC	9360
GGCAGTTTGA AACTTTTTTA CTCGCCCAA AGCGTAGcTG GtACGACTTT TTCGAAGATT	9420
GTCTAAAAGC CGGTGGCGTA ATGAAACGCT TAGATTTGGC AATCAATGAT CTAGTAGGGC	9480
TATTGGATAT TCCTGATTTA ACGAAAAAAT GTCAGAAGGA AGAATGTATT TCCTTGTTTC	9540
GTACCTTTAA AAGTTACCGT TCGGGGGAGC TGTTGAAAGC TGATGAAAAG GATGGTATGG	9600
GGAATACCTT ATACATTGGC AGTCTCAAAA GTGAAGTTA TTTTGTGTTG TACGAAAAAG	9660
ATTATGAACA ATATATCAAG CTAGGGATTG CATTAGACAA AACTGAAACG AAAAATCGCT	9720
TTGAGATTTC ACTAAAAAAT GACCGTGCCT ATCATGCAAT TCAAGATTG TTAAAGGGTC	9780
GTAGCATTGA AAGTACCACG TTTTCCATTA TCAATCGTTA CTTGCGATTG GCGGATAAAG	9840
TAGAAGGCAA AAGGCGAACT AACTGGCCAT TAAATGAACA ATGGGGGCGT TTTATTGGAC	9900
GGAATCGGAA AGAGATACAA CTGACTTCTG AGCCGAAACC GTATACCATT GAACGAACCT	9960
TAAATTGGCT AGGTAGACAA GTCGCCCTA CGTGGAAGAT GGCCAAGGAG TTGGATCGGT	10020
TGAATCAGAC AACCTATATC CAGGATATGG TACAGAATGC GCGCTTATCA GACCGGCATA	10080
AAAAGATTTT GGAACAACAA AGTATGGCAA TCGAAAACCT GATTATATGA AAAGAGGAAT	10140

TGTGGTAGGA	AGTGTAAG	ATATGTTATT	GGCATTGGT	GGCAGCATGC	TTAGTGTTAC	10200
CGTGATGTGC	TTGATGCATG	CAAGTGTAGC	GGCGAACCAA	GCAATGGAAA	AGAAAGTCTA	10260
GTGGTAATAA	GTCAAGATAA	TTTTTCAAAA	GATTTTTTAA	GGTAGTTAAT	TTTTGAAGCG	10320
CACTGAACATA	AACCCGACCA	AAATTAATAA	TTTTTTCGGG	TACTTCCAAG	GATTAGCGTC	10380
AATAACATAC	TGTTATTTAT	CTATCACGCT	CCTTGGTGGC	ATATAGAGTT	GATGGTTACG	10440
TAGTAGCGTA	TCCACCAGAC	GCGTAAATTT	TCTTGCGGTT	AGTACGATGG	CTCGTTGTG	10500
TTTATGCTTG	GGTGTTCAG	CGTATTTTTT	CTTATAATAC	GCATGGTACT	CAGGCTCGTA	10560
TCTCCTTACG	GAGTTGGTGG	CTTCAACTAG	ATAATAACGG	AGATAACGAT	TGCCTCGTTT	10620
GGCCATAGGA	GTATTTTCAT	ACTCAGAGTT	ACCTGATTGG	TTCTGTCGCC	AATATAAACC	10680
GGCATATTTA	GCCAATTTGC	TTTCATCTGA	AAAGCGATTG	ATAGGGCCAA	TTTCAGCGAG	10740
TATACCCGCA	GCATAAACAG	GACCGATTCC	TGGAATACTG	GTTAAACATT	GATATTCAGG	10800
TAGCACAACG	ACTAAATCGT	CAATGGCTTT	GTCTAAGTCT	TTAATTAAC	TTTCAAGACT	10860
ACGAATTTCT	TTGGCTAGAA	CACCCAAAAC	AATATCTACA	GAGTCCTGAT	GGACTTTTCC	10920
TAACCGGTAG	GAACCTACGA	TGGCTTTAGA	AATTGCCTTT	GCAATTCCtT	CGGGaTTTTT	10980
AAAACGATAC	TTCCCTAGTT	TTTGAATTAG	GTTGCGAAAC	TCTTCTAATG	ACAGTTCGGC	11040
AAGTTGGTCC	ATTGTGTAAT	CTTCCGTCAT	TAGAGTTACT	AATGTGGcAC	TTAAAACGGA	11100
TCCACCTTCT	GCTTTTAGTT	CTTTAGAAAG	CGTGTTCAT	TTATAATAAA	TGTTTTCGAT	11160
AAAGTGTGTC	TTGTACGAA	CCAGCTGTTT	AATTAGCTGA	TAACGCGTTC	TAGTTAAGTG	11220
TTGGAGGGCA	AGGTATTCTT	CCTCTTTAAT	GATGGAGTTG	ACTGCGCTTT	GGATGCGAAA	11280
ATAGTCAGCG	ATATAAAAAG	CATCAATTTG	ATCGTTTTTG	TTTTCTTCAA	AAATATCACG	11340
ATACTTTTTA	ATTTTATTTG	GTGTTCAC	TGAAACCATA	AGGTTCAATT	GATTTAACTC	11400
AAGATCTTCT	TTAAAAACA	TAGCTGGATG	AAAGCTATAT	AAAGATGTGG	CCTCCATTCC	11460
AATCACTAAG	CTCTCTATCT	CAAATTTTTG	AGAAAATTCG	AGAATGTATT	CTTTGATCTG	11520
GCTTGCGCCA	AGTTGACTAT	TTCCATAAGA	TGCTTCTTTT	AAAACAGAGC	AAGTTGAATC	11580
GTCTGTCATA	AAGCAAGCAT	CTAATTTTTT	AGAACTAACG	TCTAAACCGA	CAAATAGTTT	11640
CATGTGGAAG	GACCTCCTTT	CATTGATTAT	TCGGAATTTG	CTTGGGTACT	GTGGGATATC	11700
CCAGAATCAC	GTTGTCTAAC	GACAGCCTCG	CATAATAGAG	TTTCATTCAT	TAAACCTCTA	11760
AGCCACCCAA	TCTcTACTAG	GaTTGGTAGA	GGTCAATGGa	CTCAGCTAGT	GTGTGGaGG	11820
TTATACCAAC	GGTTCGGGTC	ACAGACTTTC	TGTCGCAGTA	AAACTGCAGA	AGGTGATAGC	11880
GTCTTCCCGT	AAGTCCTTTC	CAAGACTTAT	TATTCAGGAA	TATCCTACAG	TATCCAAGTA	11940
GTAATCGTTA	TTTTATTTAA	ACATATTAGG	AATAGCCCCT	AAGGGmGwTT	GAAGGAATGA	12000
AAGTTTCTTC	AAAATATATA	TTACGAGGAA	AGTGACAAGA	TGAATTCGG	ACAAAATTTG	12060
TATCAATGGT	TTCTAACCAA	TGCCCAATCC	TTAGTCTTAC	TGGCAATTGT	GGTCATTGGT	12120

TTGTTTTTAG GATTTAAGCG GGAGTTTTCT AAATTGaTTG GCTcCTGGTG ATTGCCTTGA	12180
TTGCGGTAGG TCTGGTCTTT AATGCTTctG GTGTAAAAGA TGTCyTACTC aATTTATTCa	12240
ATCgAATCGT CGGTGCGTAA GGAGGAGTCA CGATGGAACa GATGCGTGTT TATATTGCGa	12300
ACTTAGGTCG ATATAACGAA GGCGaATTAG TTGGTGcTTG gTTTACACCA CCTGTTGATT	12360
TTGATGmGaT AAAAGAmCGA ATTGGATTAA ATAATGAATA TGAGGAATAT GCGATTCATG	12420
ATTATGAACT TCCTTTTGAG ATTGATGAGT ATACACCAAT TGAAGAGATT AATCGGTTGT	12480
GTGGTTTAGC GGAAGAGCTT GAAGGAACGC CGATTGGTGA GGTGCTTCA GAGATTCAAC	12540
ACGCTTTCTT CAATTCATTT GAAGAAATGG TGGAGCATCT GGATAATATT GTTTATTATC	12600
CAGATTGTAA TGACATGAGT GATTTAGCGT ATTACTTGAT TGAGGAAGCA GGCGCTTTAG	12660
GTGAAGTTCC TACTCATTTA CAAAATTATA TTGACTATGA GGCTTATGGT CGAGATTTAG	12720
AAATTGAAGG GAATTATTTG GTTACGAGTC GTGGGTCTT TGAGCTATGT GAGTGAGAGT	12780
ATATTGAATA CTGCCCTTAA AGATACTACA ATAGAAAATA TCGAGTAGTA TTTCAAAGGA	12840
GTGGAAGTGG TGAATATTGA TGAGGCAAAA CAGCTTGCAA ATTATTTTTC AGACGGATCT	12900
CCTGGATATA TATTGGAATT CAATAATCTT GAATGGGACC TATATACTAA GAAATAGTT	12960
GGATTTGGAA TTCAAGAAAG ATTTGGTGGC TCTAAAGGAA ATTCCTTCAA AGAATTTCTT	13020
GTCTCAGAGG ATATTAGTGA TCACCTAAAG TTCAAATTAA TATACTCATa AATGaAACTA	13080
TGAAGCAAAT gCTTCTATTC GCCAATTTGa AAATCCAAGT TCaTATAAGT CaCCTAAAAk	13140
GAATCCACTT GtGACTCyAw TGaTTgAAGA GGCaTATCaT AAGaAAGATA GTGCaAGATT	13200
GCTtTAATGGT ACTAAGTTTT CGAAAGAAGT GAATTTAGAG TATATTCAAC AACTTCCTAC	13260
TCGTATAAGT AAGGATATTC AAAATAAAGA TTTCGATTCT GTTTTAACAA AATCAAACAC	13320
TATGATTGAT GAAGCTTGA AATATATCAT TGAGAACACA AAAGGAAAA GCATTGATAA	13380
TAATACTGTA AGGTCTAAAG ACTTAAGAAA TCAAGCCTTT GAAAGACTAA ATATAAAAAT	13440
TGACAAAGAT ATGGATAmAa GAATAAAaGC TTTAGTTGGT GCTCTAAACA CTTTGyCAGA	13500
CAAAATTCTy GAGATGAGGA ATAGTCAAGG TGATGCCCAT GCTCAGGGCT CCAGGAGGAT	13560
TGTAATTAAT GAAGAAGAAG CTATCCTTAC TGCGAACCA GCAATGATT TATGTGAGTA	13620
TCTATTTCGA AAGTATGAAA AGAGAAAAAG GTTCAGTGAT GCAAAAGAGA GTTGATTtTT	13680
ATCAGCTCTC ATTTTCTATC TTTyCAATTT TTGCGAATAG CTTATAAGA ATACTAGTTt	13740
GTTCTATAAC AAATTCTACT TCTAGGTCAG TGAATGTTTG ACTATGTTTA ACATTTTCAT	13800
TTTGATAGCT TATAAGTCct TTAATAATTA AACGGACCA ATTGATTATT TCTTTTGGAA	13860
TGCCCtTTTT TCTCAATAAG GGCAGGAGTT CTCCTTCGAT AGGCTGGTTT TCTAAAGACT	13920
TTGAGTTTTT TAAATATCC TGCATAAAAA ACTCAAAGGC CAGCCTGATA CTGTCTAAAG	13980
TATAACGGTT ATAGATGCTG TTGTCGTAAG ACTGTATGGC TTTTCAAAT TCTTCTTTTG	14040
CTCTTGAGGA AGCAGTCAAC CAAGTGTTAG TTTCATTGAT GATATTTCTA TCTATTTTAT	14100

CATTAGAATA	AACTGAACCA	TAATTTTTTA	GTAACAGTTT	TAAAATCTTA	GCAACGTCTT	14160
TATTTGCTGG	ATTTTGAGCA	AGCACACTTT	CACAAAGCTC	TTTTATAATT	TTAAATTGCT	14220
GTTCTTCTGT	GAAAGATTGT	AAATTATTTT	GAATCGCGGT	AGCCTTAGAA	GTAAAGTTCTG	14280
gAAAGGGAGT	CGAAATAAAG	GGAACAGAAg	AGTTAAATTC	TGCTCCATAT	TCAGCAAATA	14340
AAGTTGCAAT	CTCAGTCCCC	GAGTAAACTT	CGTTAAGTGT	TCCACTTGCC	ATTGTTATTA	14400
AGTTTTGCAT	ATATACCTCC	ATAAAAAATT	AAGCATGATG	TTTATATTAT	ATTAGAnAGT	14460
GAGCCGATTG	AATATGAAAA	AAATAAAAAG	CTACACCAGT	ATCTGGGCAG	TTGAAAAGGT	14520
CATTTATGCA	ATCAATGACT	TTCAGCTACC	TTTTCCTGTG	ACCTTTAATC	AGATGGCTTG	14580
GTTTGTTTTG	TCCTTGCTGT	TTGTCAFCGT	TTTTGCCCAT	GTGCCTCCGC	TATCTATGAT	14640
AGAGGGGGCC	TTTTTAAAT	ACTTAGGCAT	ACCGGTGGTG	GTCACTTGGT	TCATGTCGCA	14700
AAAAACCTTT	GATGGTAAGA	AACCACTGGG	ATTTTTACGA	TCCTTTCTCA	GTTATCACTT	14760
ACGGTTTAAG	GTCACCTTTG	CAGGAAAGAA	GGTCAAAGAA	CAGAAGAAAC	GATGGGATGA	14820
ACCCATCACT	CTAGTAAGGA	GTGTGAACTA	TGTACCCGAT	TAAGTATGTT	GAAAATAATC	14880
TTGTCTTTAA	tCAAGAGGGa	GAATGTTTTG	CCTACTAwGA	ACTTGTGCCT	TAwAACTATT	14940
CCTTCTTATC	GCCTGAACAG	AAGTATCAAG	TGCACGATAA	CTTTCGTCAA	TTGATTGCTC	15000
AGAATCGTGA	AGGCAAGATC	CATGCCTTAC	AAATTGCCAC	AGAAAGTAGT	ATTCGTGCCA	15060
CGCAAGAACG	ATCAAAAAAG	GAAATTACAG	GCAGACTCAA	AGAAGTCGCG	AAACAACGAA	15120
TTGACCTTCA	AACAGATGCA	TTAGTATCTA	TGATTGGCGA	TAGTCAGATT	GATTATCGCT	15180
TTTTTATTGG	CTTTAAACTA	ATTGCAACAG	ATGAAGAGGT	CAATTTAAAG	AGTCTCAAGA	15240
AATCTTTCCT	TTCTGGGCTT	CAAGAATTG	TCTATGGGGT	GAACCATCAT	TTGATGGGTG	15300
ATTTTGTTTC	CTTATCCAAT	GAAGAAATTC	GCCGATATAC	CAAGCTTGAA	AAACTAATGG	15360
AAAGTAAATT	GGCTAGAAGA	TTCAAAGTCA	GACGGGTAC	ACCGAGTGAT	CTGACGTATT	15420
TAATTGAACA	TATTTATGGC	GAAAAAGGCA	CACCGTTTGA	AGAATATGAA	TTTCAGCTAC	15480
CAAAGAAAAA	GCTAAAATCA	GAAACGCTGG	TGAAACGTTA	TGATTTATTA	CATCCTAGTC	15540
GTTGCTTGAT	TGAAGAAAAT	CCTCGGTACT	TACGCATGGA	ACATGAAAAT	CmTGAATCGT	15600
ATGTAGCCTA	TTTGACAATC	AATACGATTG	TAGGGGAGAT	GGAGTTTCCT	TCTTCAGAAC	15660
TTTTTTATTA	CCAGCAACAA	CAATTCACCT	TTCCATTGA	TACTTCAATG	AATGTGGAGA	15720
TTGTGACCAA	TAAAAAGCT	CTTGCCACTG	TTCGCAATAA	GAAAAAGAA	CTGAAAGATT	15780
TAGACAACCA	TGCCTACCAA	TCCGATAATG	AGACAAATTC	CAATGTCTTA	GATGCTTTGG	15840
ATTCAGTAAA	TGAGTTGGAA	ACTACCTTGG	ATCAATCCAA	GGAGTCGATG	TACAAATTAA	15900
GTTATGTAGT	TCGAGTGAGT	GCGGAATCCG	TGGATGAATT	AAAACGCCGT	TGTGATGAAG	15960
TGTTCGATTT	TTATGATGAT	ACTAATGTAA	AACTAGTCCG	ACCCTTTGGC	GATATGATGG	16020
GGCTACATGA	AGAATCTTTG	CCCTCAAGCA	AACGCTATAT	GAACGATTAT	ATCCAATATG	16080



TTACTTCTGA CTTTTTAGCG GGACTTGGTT TTGGGGCTAC TCAAAAGCTA GGGGAATTAG	16140
AAGGCATCTA CTTTGGCTAC AATGTAGATA CGGGGCGGAA TGTTTACCTG AAACCGGCAT	16200
TAGCTTCACA AGGGGTCAAG GGTTCAGTTA CCAATGCTTT GGCTGCTGCG TTCCTTGGTT	16260
CACTCGGTGG CGGAAAATCC TTTAGTAACA ATCTCTTGGT CTACTATGCT GTTCTTTTTG	16320
GTGGCCAAGC TGTGATTGTC GATCCAAAAG GAGAACGAGG TGGTTGGAAA GAAAACCTGC	16380
CAGAAATTGC GGATGAAATC AATATTCTCA ACTTGACTAG TAAACCAGAA AATCAAGGCT	16440
TACTTGATCC TTATGTGATT ATGAAGAAAA AGAAAGATTC AGAAAGTTTA GCAATTGATA	16500
TCCTGACTTT TTTAACAGGA ATTTCCAGTC GGGATGGGGA GAAGTTTCCA GTATTGCGCA	16560
AAGCCATTTC GACGGTTACC CAAAGTGATC AACGAGGATT GCTGCAAGTC ATTAATGAGC	16620
TTCGAAAGGA GAACTCTCCA GTCGCGGAAA ATATCGCCGA TCATATCGAG TCCTTTACCG	16680
ATTATGATTT TGCCCATTTG CTTTTCTCAG ATGGTTCGAT TACCCAATCA ATTAGTTTAG	16740
ATAAGCAGTT GAATATCATT CAAGTAGCGG ATTTGGTTTT ACCAGATGCA GAAACCACTT	16800
TTGAAGAATA TACCACGATG GAGTTGCTAA GTGTCTCTAT GCTAATCGTC ATTTCCACGT	16860
TTGCCTTGGA TTTTATCCAT AGTAATCGAG AAATCTTCAA GATTGTGGAC CTGGATGAGG	16920
CTTGGAGTTT TCTTCAAGTG GCTCAAGGAA AGACCCTTTC AAATAAGCTC GTTCGTGCTG	16980
GTCGGGCCAT GAATGCCGGT GTTTACTTTG TCACGCAAAA TAGTGACGAT TTAGTGATG	17040
AACGCTTGAA GAATAATATC GGATTGAAAT TTGCTTTTCG TAGTACCGAT ATTCATGAAA	17100
TCAAGAAGAC TTTGGAATTT TTTGGGTTGG ATCAAGAAGA TGAAAGTAAT CAAAAACGCC	17160
TACGAGATTT GGAAAACGGC CAATGCTTGA TGCAGGATTT GTATGGTCGG GTAGGCGTGG	17220
TTCAGGTACA TCCAGTCTTT GAAGAACTCT TCCATGCCTT TGATACCCGT CCACCTGTTC	17280
AAGAGAGAAT GGAGTGAGGG GATGAAAAAG AAACTTTTTC GGATTGTGGG AATCGTTTTT	17340
ATTAGCATAC TCACTCTGAT AGTTCTATTG AGTTTAGTTG GTACGGTAGC CGAGGCGACA	17400
GGGCTAGTTG ATGATACAGT TGAATCAGGA AATCTATATT CCAAATATTC ACTGAACAAC	17460
TATCAACTAG ATTTCTTTGT GGATAGCTCT TGGGATTGGT TGCCTTGGAA TTGGGGTGAT	17520
GGCTTAGGGA AAAGCGTGAT GTATGGTCTT TATGCCATTA CTAATTTTCAT TTGGACGGTG	17580
AGTTTGTATT TGTCGAACGC CACGGGTAT GTGGTTCAAG AAGCCTATAA ATTAGATTTT	17640
ATTCGGATA CGGCTGAAAG TATTGGAAAG AATATCCAAA CATTGGCTGA TATTACTGAA	17700
AATGGACTCC AGACCTCCGG ATTTTATTTT GGGTTTCTAT TATTGATGAT TTTAGCTCTG	17760
GGGGTGTATG TGGCGTATAC CGGATTATTA AAACGAGAAA CCACGAAAGC AGTTCGAGCC	17820
GTGATTAATT TTGTGATGAT TTTTCTACTA TCGGGTTCGT TCATTGCTTA TCGCCTACC	17880
TATATCACAA AAATCAATGA CTTTAGCTCA GATGTTAGTG AAGCGGCCCT TACCCTTGGA	17940
ACCAGATCG TTGTACCCAA TTCAGAAAGC CAAGGAAAAG ATAGTGTTGA TTTGATTCGG	18000
GATAGTTTGT TTTGATTCA AGTCCAACAA CCTTGGTTAC TACTGCAATT TGATGATTCC	18060

AACATAGAAG	AAATTGGTGA	AGACCGGGTC	AACAAGATTC	TATCTGTGAG	TCCAGATGAA	18120
AACAAAGGCA	AAGATCGGGA	AGAAGCGGTC	AAGGCAGAGA	TTGAAGACAA	TGACAATGCG	18180
AATCTCAGTA	TTACCAAGAC	CATGAACCGT	TTAGGGATCG	TGGTCTTTTT	GGTATTGTTC	18240
AATATTGGCA	TTTCCTTCTT	TGTTTTTCTT	CTAACAGGGA	TTATGTTGTT	CTCGCAAATT	18300
CTCTTCATTA	TCTTTGCGAT	GTTTCTGCCA	ATTAGTTTCT	TATTGTCCAT	GTTGCCGACT	18360
TATGAGAGTT	TGGGCAAGAT	GGCGATTATT	CGTTTATTCA	ATACGATTAT	GATGCGAGCG	18420
GGTGTACCT	TGGTGATTAC	GACTGCATTC	AGTATTCTTA	CGATGTTTTT	CAATATTTTCG	18480
GCTACCTATC	CCTTCTTTAT	GGTAGCTTTT	CTACAAATTG	TGACCTTTGC	CGGTATCTAT	18540
TTCAAATTGG	GTGACATTAT	GAGTATGTTC	AACCTTCAGA	GTAATGACAG	TCAGTCAATG	18600
GGCAGACGGG	TCATGCGAAG	ACCGCAAATG	TTGATGAACC	GTAAATTGCG	GCAACTCAAT	18660
CGAAATGTGG	GCCGAACTCT	GGCTTTTGGC	GGTGCTGCAG	CAGTygGaaT	GGCAAAATGC	18720
ATTACTCTGA	TTTATGGTAC	AATTTTCCCA	TTAATCCGCC	GTATGTACCA	ACGGAAAACC	18780
CAACAGGTAT	TTACAAACGG	ACATTTGCTA	TAGATGAAAC	ATTTACAGAT	AAAAAAATCA	18840
TTTTACGTTT	TTGTGGCGTA	GATTACAGTT	ATCATGTTTG	GGTGAACGGC	CATGAAGTTG	18900
GCTACAGTAA	AGGTGCACGA	AATGAAGCAG	AATTTGACAT	TACGCCTTAC	GCTAAAATTG	18960
GTGAAACAAA	TGATTTAACG	GTGCGAGTCT	ATCAATGGTC	AGATGGTACG	TATTTAGAGG	19020
ATCAAGATAT	GTGGTGGTTA	AGTGGTATTT	TTCTGTATGT	TGAATTACTC	GGTGTACCAG	19080
AAAATGGGCT	GGAAGATTTC	TTTATTATTT	CCGATTTAGA	TGATTCGTAT	CAAATGGGC	19140
ACTTGGAAT	TACGGGTAAA	TTTTGGCAAG	ACAAAGGACA	ACAAGTTCAA	TTGGAGTTGA	19200
TGGATCAGCA	GGGAAAAACG	GTTCTTAAAG	AGACTGTCGC	AGGAAATCAA	GGTACGGTTG	19260
AATTTTCAGC	AAGTTTGCCG	TCTGTCACTG	CTTGGTCAGC	TGAAAAACCA	TACTTGTATC	19320
AACTATTTAT	CACTGTTTTT	TCTGAAGGGG	AAGTCGTCGA	AGTTATTCCG	CAAAAAGTTG	19380
GTTTCAGAAA	TATTCATGTA	TCTGGTGAGA	CTTCTTAGT	CAATGGTGTG	GCCATTAAAC	19440
TAAAAGGAAT	GAATCGTCAT	GATTATAATC	CGAAAAATGG	TCGAGTGGTT	AGTCGTGAAG	19500
AAATTGAAAA	AGATATTCGT	TTAATGAAAC	AATTTAATAT	TAATGCCATT	CGAACCAGTC	19560
ATTATCCAGC	ATCTGCTTAT	TTCTATGATT	TATGCGATGA	ATATGGAATG	TATGTGATTG	19620
ATGAAACTGA	TTTGAATGT	CACGGCTTTG	AATTAACCGG	AGAATATGAT	TGGATTAGCA	19680
ACGATCCAGA	ATGGGAAACG	GCTTATGTTT	CTCGGATGGT	ACGGATGATT	CAACGGGATA	19740
AAAATCATCC	ATCAATTCTC	TTTTGGTCGT	TAGGGAATGA	ATCTGCGTTT	GGTCATAACT	19800
TTATTGAAAT	GGCCCGTATT	GCAAAAGAAA	TGGATCCAAC	ACGTTTGGTT	CATTATGAAG	19860
GTGATTTTGA	GGCAGAGGTC	ACGGATGTCT	ATTCTACAAT	GTATACGTGG	TTGGAACACC	19920
CAACTAGAGA	ATTACTGATG	AATACGATTA	TTGAAAATTC	AAAAAAACCG	CATATTTTAT	19980
GTGAATATTG	TCATGCGATG	GGTAATGGCC	CTGGGAATTT	AAAAGAATAC	CAAGAGCTAT	20040

TTTATGCTCA	TGATAAATTA	CAAGGTGGCT	TTATTTGGGA	ATGGTTTGAC	CATGGCATTG	20100
AATCTGTAAC	AGATAACGGT	GAAGTTTATT	ATCGTTATGG	CGGCGATTTT	GGCGATGATC	20160
CAAGTAACAA	AGATTTTTGT	ATTGATGGGA	TGTTGATGCC	TGACCGTACG	CCATCTCCTA	20220
GCTTGTATGA	ATACAAAAAA	GTCATTGAAC	CCATTACAAC	TAGCGCTATT	GACGTTCTGT	20280
CAGGTGAATT	CTCATTATTG	AGTCGGTTTG	ACTTTGAAAA	CTTGGCTATC	TTTAAGCTGG	20340
TTTATACAAT	TACTGAAGAT	CAAACAGTGA	TTCAATCAGG	GACAGTAGCA	GTGCCAGCAA	20400
TTGCTGCTAG	AGCAGAAGGA	AGACTGCACT	TACCTTATCA	TTTAGATTTT	CCTAAAAAGG	20460
CAGGAGCCGC	CTATTATTTA	ACTCTTTTCAT	ATCAATTGAA	AGAAACGACT	GCCTATGCAT	20520
CTGCTGGTCA	TGAATTAGCT	ACCGCACAGT	TTGAATTGCC	AATTGCAACA	CCAGGGATTG	20580
AGATCACACC	AGTTGGCTCG	CTAATGGCAA	AGGAAATTGG	CCCTCATTTA	TACATTGAAG	20640
GACCAAACCT	TTCAATTAAT	TTTGATAAAG	TGAAGGGCGC	CTTAACGAAT	GTGACACGAG	20700
ATGGTAAAAA	ATTATTGCAT	AAAGGGCCGA	AATTTACATT	TTGGCGGGCA	CCAATCAGTA	20760
ATGATATGGA	AATTATTGAT	GAAATGAAGA	AGAAATATTT	CTTACATTTA	GAGCATGAAA	20820
TTGTCCGTTT	TTTTGAGTGG	AAAAAAGTTG	ATGATTTTAT	CCAAGTCATC	GTAAAAACAA	20880
TTAACGGAAC	GACTAACAGT	GCTTGGCATT	ATCAATGTAC	CTATCAGTAT	TTAATTGCAC	20940
CAAATGGCGA	AATTTTCTTC	GACTTAAAAG	GTAGTCCTGC	TGGTAAAATT	GAAAATGCGC	21000
CAGATATGTT	ACCACGTTTG	GGTGTGACTC	TCCATTTAGA	CAAGTCTCTA	TCAGAAGTAA	21060
AATATTTTCGG	TAAAGGGCCG	AGAGAAAAC	ATGTAGATTC	ACAAGAAGCT	GGTTTACTTG	21120
GCGTTTATGA	TGCTACAGTG	GCGGAAATGT	TCACTAATTA	TGTTGTTCCA	CAAGCCAATG	21180
GTAACCACAT	GGCGACTAAA	TGGTCCGCAT	TCACGGATGA	TCGCGGCCAA	GGGGTAGTAG	21240
CAACAGCGGC	TGATTCATAT	AATTTTAGTG	TTTCTTATTT	TGAAGAACAA	GCGCTAGATG	21300
TTGCTAAGCA	TACCAATGAA	TTACAAGAAA	GTGAGTATGT	TGTTTTAAAT	ATCGATTACA	21360
AACAAAACGC	CCTAGGTAGC	TACTCTTGCG	GCCAGTGGCA	ATTAGAAAAA	TACCGAACTA	21420
CGTTTGAAGA	ATTCCAATTG	GCATTCCGCT	TAACACCTTT	TAATAATAAA	GAGATTCAAG	21480
CAGCGGATGT	CGCACATGAA	CGAGTAAAC	GACCAACAAT	TTCTTAAATA	AAAGCTCAGA	21540
ACAGTGCACT	TCGGTCTCCT	GTTCTGAGTT	TTTTTAATGA	AATTATCAGA	AATTATTGCT	21600
TTTTTTTTCAT	GGAAATAAGC	AAACTATTTT	GTTATCTAAT	AAAAATAGTT	CATAATAGTA	21660
ACATAGCGAA	TTATTCGCTA	GATAAAGCAT	TTAGGAGGAA	TTATTTTATG	CATACTTTAT	21720
GGGTTTTAAT	CGTTGGTGCT	GTTATCGGTG	CTAtCgcTGG	CGCGCTTACA	AGCAAAGGTC	21780
AATCAATGGG	TTGGATTGCG	AATATCGTTG	CCGGTTTAGT	TGGTTCATTT	ATCGGAGAAA	21840
AAATCTTAGG	TAGCTGGGGT	CCTCACGTAG	CAGGTATGGC	AATTGTGCCA	TCAATTATTG	21900
GGGCAGTAAT	TTTAGTTGCT	GTGGTTTCAT	TCTTCTTCGG	ACGAAAAAAT	AGCTAATTAA	21960
GTAAATAATT	GTTTAATTAG	TTAAAGGAGG	TCGGGACAGA	ACCGTTTAGC	TCCGAGAAAT	22020

AAGAAGAAAT	TTCCGAAAAT	TGCTTTTAA	TTTTTGATA	ATTTCCGCTT	ATTTCCGAAG	22080
GAGTTGGTTC	TGTTCCCGCC	GTTCATCAGT	TTTTGAGCGT	GGAGCAAAAA	TCCAAAGTGA	22140
TTTTTGTC	ACGCTCCTTT	TTTTATTTTC	TCGTAAATTT	TACGACAAC	TCACAACGAG	22200
CGGTTTGTGG	AAACATATCA	ACAGATTGTA	AATAGTCTAC	TTGATATACT	TTGGCTAATT	22260
GAACATAATC	ACGTGCTAAC	GTGGATACAT	TACAAGAAAT	ATACACCATT	TTCTTAGGTG	22320
GTTGTTTCAA	AAGGGCCGTT	AATAATTTAC	GATCCAAACC	AGTCCGTGGG	GGATCAACTA	22380
CAATGGCATC	AGGTTTAAAG	CCTTCTTTGA	ACCATTTAGG	TAATAAGTCT	TCAGCTGTTC	22440
CAACAGCATA	ATGTGTGTTC	GTAACACCAA	GTCGTTTTGC	ATTGAATCTA	GCATCGTCAA	22500
TTGCGGCGGG	AATGGTATCC	ATACCACGAA	CTTGATGCGC	TTGTTTCGCT	AAACTTAAGC	22560
CAATGGTCCC	GACGCCACAA	TAAGCATCAA	CAACAGTTTC	GTTTGTTTGT	AAATCTAAGG	22620
CCTTGATTCC	TTCATTATAT	AAACTTCCG	TTTGTTCTGG	ATTCAATTGG	AAAAATGCTC	22680
GCGGCATAA	ATCAAAGACA	ACGTCGTTAA	TGTGTTCTTC	AATGCTTTCT	TTGCCCCATA	22740
AGTGAAGGGT	ATCATCACCC	ATTACTAGAG	AGGTTTTTTT	ATTTTGAC	TTTGCATAA	22800
TAGAAACAAC	TTCTGGTAAT	TGTTCAATTA	TTGCGCGAAC	CATTTTTTCT	TTTGTGGAA	22860
ATTTCTGCGA	TTGTGTAATG	AACACTACTT	GGACTTCGCC	AGTTTGATG	CCCACTCGAA	22920
CCATTAATGT	TCGAAAGATA	CCACTATTTT	TACGTTTCATC	ATAAATTGGT	AATTGAAAGT	22980
CATTCAATAA	ATCAACTAAT	GTATTCATCA	CCTTGTTTGT	CGCAGGCTGT	TGGACCAAAC	23040
AGTTATCAAT	CGGTACTAAT	TGATGTGAAT	CAGCTTGATA	GAGACCAGCT	TCTACTTGGC	23100
CATCGATTTG	GCGTAATTGG	AATTGGGCTT	TATTTGATA	ATTCCAAGGA	TTATCCATCC	23160
CAATTGTTTT	TCGTAGTTCA	TAGTTTTGAT	AATTGGCTGG	TTAAATTTG	TTAAGCGCCT	23220
GCTTTAATAA	ATCTTTTTTA	AAGTCTAATT	GTGCTTTGTA	GGCTAAATGT	TGTAGTTGGC	23280
AGCCGCCACA	TGTTTCATAG	ACAGGACAAG	GAGGGACCAC	GCGATCTTTA	GCAGCTTTTT	23340
TGACAGATTG	AAGGGTTCCT	TCAGCAAATT	TTGGTGTAAC	ATTTGTGATC	GTGGCAGTCA	23400
CTTCTTCTTT	CGGTAGCGCC	CCAGGAACGA	AAATAATTAA	TTTTTTATAA	TAGCCGATAC	23460
CTTCGCCATT	AATTCCTAAT	CGTTTAATTT	TTAAAGAAAT	TTGTTGGCCT	TTTTTTACTA	23520
GTTGTTGGGT	CATATTCTTT	CATCCTTTAG	TCATCTCACA	TTTTTCACTT	CCTCCCATTT	23580
TAGCACATTT	TTATTTATTT	GGCTTCCAAG	GATTCTATGG	TATAGTAGAA	TATATGTTCT	23640
GATTTAGAGG	GGCAGGAGAA	TAAGATGACA	ACAATCACAC	GAATTAGTAA	AGATAAAGGT	23700
GAATTTTATC	TACTTTGGCT	ATCTTCAGGC	GAAAACTGC	GTGTCTCGGA	AGATATTTTA	23760
GTACGGCAGC	GTCTGCTAAA	AGGCCAAGAG	TTATCAGATA	CTTTAATAGA	AGAAATAAAA	23820
AAAGCCAGTT	CTTATGATGT	CGGGTTACAA	ATGGCGATGA	ATTATTTAAG	CTATCAATTA	23880
CGTTCAAAGA	AAGAGATTTT	TACTTATTTA	AAAGAGAAAG	AAATTGTACC	TGAAGATCGA	23940
GTAAAAATCG	TTCAACGTTT	AGAAGAGCTG	CGGTTGTTGG	ATGATGCAAT	TTTAGTGAA	24000

AGTTACGTAC	GGACAGCTAT	GCGAACCAGT	GATAAAGGAC	CACGAAATGT	GGCACAACAA	24060
TTAAAGCAAA	AAGGTATTAG	TGAAGAAGAT	ATTCAGCATG	GTTTGACCTT	TTATACGTTA	24120
GATGAACAAC	TTAATGTTGC	CACAGCAaCT	GCAGAAAAAG	CCATGAAACG	CTATCGAACT	24180
AAAAGTTTTa	AAGATGCACT	TCaGAAAAATG	CGTCTGCATC	TGATGCAAAA	GGGCTTTACA	24240
AATGAAATCA	TTGATCTAGC	TTTGGAACT	TTAGCTTTTG	AAAAGGACGA	AGAACAAGAA	24300
CAACAAGCGT	TAGACAAGGA	AGGCGAACGT	TTATGGCGTG	CCAATCAGCG	ATTTGATTTT	24360
TCTAAAAAAG	TCCAAAAGT	CAAAACAAAGC	TTATTCCAAA	AAGGGTTTGA	CTATGATTTG	24420
ATTCAGCAAT	TTATTAGCGA	AAAGGAAGTA	GAACATGACG	AATAAAGAGC	CTTGGGAGGC	24480
GTGGAGTCCC	GCAAAAGTTA	GTTCAATTCCA	GGAAGATTTT	TTAGCTTGGT	ATGAACGAGA	24540
AAAACGCAAT	TTACCTTGCG	GAGCGAATAC	AGATGCATAT	CGTATTTGGA	TTTCTGAAAT	24600
TATGCTACAA	CAAACCTCGC	TAGATACAGT	CATTGATTAT	TTTTATCGAT	TTATGGAATG	24660
GTTCGACG	ATTCAAGATT	TAGCGGAAGC	GCCAGATGAT	AAGTTGTTGA	AAGCTTGGGA	24720
AGGGTTAGGT	TaCTATTAC	GAGCGCGTAA	TTTAAAAGTG	GCAGCGCAAC	AGATTGTTTC	24780
AGAATTTGGT	GGGAAAATGC	CTGACACAAT	CGAAGATATT	CGGAGTTTAA	AAGGAATCGG	24840
CCCCTATACG	GCTGGTGCGA	TTGGCAGTAT	TGCCTTTAAT	CTCCGCGAAC	CAGCAATTGA	24900
CGGTAACGTC	ATGCGGGTAG	TGAGCCGTTT	ATTTGAAATT	GATGCAGATA	TTGCAAAAGC	24960
GAGTAGTCGT	AAAGTATTTG	AAGCGGCGAT	GCTTAAAATC	ATCGATCGTG	AGCGTCCAGG	25020
GGATTTTAAT	CAAGCTTTAA	TGGATCTAGG	CTCAGCCGTC	TGCACCCCAA	CTTCCCCTAA	25080
GTGTGAAAGT	TGTCCGCTTC	AGCAATATTG	TGCGGCCTAT	CAGGCAGATA	AAATGACCGC	25140
TTATCCAGTC	AAATCAAAAA	AAGTTAAGCC	GAAAGATGTG	TATTATGTCTG	GCACTATTAT	25200
TGAAAATAAA	AAGCAAGAAT	TTTATTAGA	ACAACGCCCT	GAAACCGGTT	TGTTAGCCAA	25260
TATGTGGCTG	TTTCCGATTG	AAGAAATTAG	CAAAAAACAA	TTTCAACAAT	TGCAAAAAC	25320
GGCGCAACCA	GCAGAAACGG	AAAAACAATT	AACTTTAGAG	TTAGAGCCTG	TAACAGAGCC	25380
TTTGGTCGCT	GAAGAACCTG	TCAACTTTTT	TACGGACTAT	GAAACAGTTG	TCTGGCAAAA	25440
ACGTACCTTA	GCGAAGTCG	TTCATATTTT	CAGTCATTTA	AAATGGCATA	TCTTAGTCTT	25500
TTATGGCCGA	AATACAGGCG	AATTAGCCAC	ACTAGAGTCG	CAGCGATGGG	TGGCAGCACA	25560
ACAATTTTCA	GACTATGTTT	TTCCAAAACC	ACAGCAAAAA	ATGGTGGAAT	TGTTCAAAAA	25620
AGAACACGAA	AAAAAATAAC	AAAAAAGAGA	GGCTTTTTTA	CAAACAAGCC	CCTCTTTTTT	25680
GTTTCCTCTA	GTTTATACCC	CTATATAAAT	TCTGCTATGT	TTAGTTTTGA	AAGGGACTGG	25740
TAAATAGAC	TTGTCTCTTG	TCTGCTTAGA	GGAGTAGTAA	TGATGGTGGA	GACTTTTTTA	25800
GAGAGAGGAA	GTACAGCCAA	TGAGTAGGAA	GCGAAAAATC	AGCTTAATTA	GTTTAGTCAT	25860
CATTTTGGTT	TTGTCTACAG	TCGGCTCAGC	ATACTTTGCT	GTAGCGGGTA	GCTATTTAAA	25920
GAAAACAATT	GATAAAGGCT	ATGTTCCCAT	AAAAAATGAT	TATAATGAAG	CGCAAAATAA	25980

AGATAGTCAA	TCGTTTTTGA	TTATGGGGCT	AGACAATACA	ATTGAACGGA	AATTAGGCAC	26040
AACTAGGACT	GATGCTATGA	TGGTGATTAC	CGTGAATAAC	AAGACGAAGA	AAATAACCTA	26100
TTTAAGTTTG	CCACGGGATA	GTTTTGTTC	AATTGATGCG	AAAAATTACC	AAGGGATGCA	26160
GCGAATTGAA	GCCGCCTATA	CCTACGATGG	ACCAACAGCT	TCTGTTAACA	CAGTTGAGAA	26220
ATTATTGAAT	ATTCCAATCA	ATCATTACGT	TGTGTTTAAC	TTTTTATCTT	TTATTAAGTT	26280
AATTGATGCG	GTTGGCGGCA	TAGATGTCAA	TGTCAAGCAG	GCGTTTGATG	GTGTCACCAA	26340
AGACGGGCCA	GGATCCATTC	ATTTTGATGC	AGGGAAACAG	CATTTAGATG	GTACGAAAGC	26400
TTTATCTTAT	GCCCCGTGAAA	GACATAGCGA	TAACGATATT	ATGCGTGGAT	TCCGACAACA	26460
AGAAATTATT	CAAGCAGTTG	AAGACAAGTT	GAAATCTGGT	CAATCAATCA	TGAAAATAAT	26520
GGACATTATT	GATTCGTTAA	ATGGAAACAT	TCAAAC TGAT	GTGGATTCCA	ATGAATTGAC	26580
TCATTTAGTC	AAAGAAGGTT	TGACTTGGAC	CAATTATGAT	AAACAACAGC	TTTCTTTTGA	26640
CTGGCGCACT	TTTAGTAATG	AAGGGCGCAG	TATGGTTGAA	CTATACCCAG	ATAGTATTGA	26700
AAATGTCCGT	CATCAATTAC	GTGTGTCTTT	AAATTTAGAA	AAGCCAGATG	AACGAGATCA	26760
AGACGGCTAT	GTCTTCATA	CGAACGGTGA	ATTTTTATAT	CAAAGTGATT	ATACCGTTCA	26820
AGATGAAGCA	GCTGAGGAAA	ACGAAATGAC	TTCCATCAAC	GGCAATACGT	ATATTGGTGT	26880
TCCTGGTAAT	ACACAGACCG	GCCCCGTGCC	ATCAGTTAAA	ACGGAAAATG	GCTTTATAAA	26940
ATAATTTATT	TTCCAAGTGT	AGAAAATCCG	AACATAACAG	ATTAAATAAT	CCGCTTACTT	27000
ATTATTCTTC	TAAGGAAAAA	ACGAAGGATT	TTAGCAAGGA	AACAGCGTAT	CATTTGATAT	27060
AAACCCAACA	TTCCCTTTAT	AATGAAAGTA	ACAAGCAAAT	TTTTGTTACT	TAGAAGGGGA	27120
AGGGAGCAAG	GAGTTATGTT	TAGTTATACA	GATACCATT	TGTCAATTAT	GCAAAGAATT	27180
GAAGTATACA	ACGAAATTTT	CAAAGCCATC	TCTAAAGAAG	TTCAAGAATA	CGGTTGTAAT	27240
CAGGCTGCAC	ATCGGAGAGG	CCGTGACACT	TATATGTTTT	GCCGAAATAA	TGTGAATCGT	27300
TTTTTTACAG	AAGAAATGAT	GTTTCGTAAA	AATTTAGAGC	CGTATTATGA	TAAAGAAGCA	27360
ACGAAAATCT	TGCTAGAAGG	ATTAGATACG	TATAAAGAAG	GAGTATTTTT	CTGGTTAGAG	27420
GCATTAAGTG	AATCTTGCCA	AGTGACAGAT	GAATTGAAAT	ATCAATTAGG	AATCCGCCAA	27480
AAAGAATCCT	CTTATTTATT	AATTAATCAA	GCTTGTAAG	AAGCCTGTGA	GGCATTCAA	27540
AGCGCTCATA	GTGTTCATAA	GATGTAAAAA	GTTAGGGATA	GCTGTCAATC	TGGCAGTTAT	27600
CCCTAACTTT	TTTATTTCAA	TATCATCTGA	TAAGCATAAC	GTTCGCCATC	GTTAACATCT	27660
AAAATCACGT	CACCTTGATA	AACAAAGCCA	GCTTTTTTAA	TCAGGTGCTG	CATAGCGAGA	27720
TTTTGTGGAT	GTGTATCGAT	TCGAATATCA	GTAGCACCTT	GCAAACGAGC	GGCTGTAACC	27780
AGTGAGCGCA	AAAGCACCAA	AGCCAGTCCT	TGACCTTGAA	AGGAACTTTT	TAGAGCAACA	27840
CAATGAATTG	CCGTATACGT	AGTTTGAGAA	TGAGGTTGCC	AAGAACCATT	AAAAATCGCT	27900
TCATAAGCAG	GATCAATCTC	CGAACATAAA	ACACCAACGC	CTGCCAAGGC	ATCCTCAACC	27960

ATTAAAATAT AGCAAGTTTG CTGGTCGATA TCTTTTGCTA AAATTTCTTG ATTCGGGCCA	28020
TCGCCGTTTT GCCATTGGGG rATGCCACTT TTTTGTAAGG TGCGGCGTCC ATCTTCAATG	28080
ATAGCAAGGA TGTCAAGGCAA ATCAGAAGGT TCTGCTTTTC TTAAGTAAAC AGTCAAAAAA	28140
TATTCCCTCC TCTTAGAAAC CTGTTGTTGT TTTCAATTGG TAAATTTTTG AAGGGCGGCC	28200
AATGCCAACG GGACGTTAC CTAACCTCTC AAAATATTGT GTCATGGCTT TTTTAAATT	28260
TGAATGGTCA ATGCTGCGAT AATCAACGCC GAGGAACCTC GCAAAAACCT TACGGGCTTC	28320
CGTAATGGTG AAATCTTTTC CTAAGACTTG AAGCACTTGC GGTTCATGTT CCATTTTATC	28380
AACAACACGA TTAAACGCCT TAATTATAAT TTCACTATGG TCGAAGGCTA ATGTATCTTT	28440
ACCAAGAGAG GCCGCTGTTT TCAAGTCTAA AGTAATTTCC ACATCTTCAT GAGATAACGT	28500
GATGTGTTGA CCATGCCGTT CAAGATTGAA CCAATGGACT TCTTTGGCAT CGTCGCCAGC	28560
AATTAAAGGT TCTTCACCAA TAAAGGCTAA ATAGCTGACT GTCACGACCC AGCCACGAGG	28620
ATCTCGATCT GGGCGACTAA AGCTGTGTAA TTGTTCAATA TTTTCTTGAG AAATCACCAC	28680
ACCAGTTTCC TCTTTTGTTT CTCTTAAAC ACTATCTTCA GTAGATTCGT TCGGATTAAC	28740
AAAACCGCCA GGAAGCGCCC ATGAGTTTCT AAAGGGTGA CCTTTACGTT GAATCAGTAA	28800
GACTTTTAAT TGATCAGCTT CCTTGTTGTA ACAAAGAAGA ACCATATCAA CGGTTAAAGA	28860
GGGTTTTTCA TATTCAGGTA ATTCTTGTTG ATGATACCAA GTTAAAACT CTGCTAAGCT	28920
AGCTTGCGCT TCATAATAAT TTTTTCTTC GGCTTTAGAA GCAAATTGTG GCACAATTGC	28980
TCCTCCTTTC AAAAAGAAAC ATGGTTTTCT TTTATTGTAT CATTAAAGTG CTGAAATACA	29040
ATATTCAAGC GTAAAAAGAA AAGCAAGCCT GGGACAAAAA TCACTTTGGA TTTTGTCCC	29100
AGGCTGTAC TATAAGAAGT TTGTTCTAGA CGTTATTTAT TGTGTTAATT CTAATAATTT	29160
ATTTTCAAAA TCTTCTCCA TATGGCCAAG TTCGATTTCG GCAGCTTGGC GTTTTCTTT	29220
TCCTTCTTGT TGAATACGTA ATGTTTCTTG GATAGTTTCA ATTAAATCAT TTTGTGTTTT	29280
TTGtAATGTT TCAATATCGA CTAGGCCACG TTCGTTTTCT TTTGCTGTTT CAATTGCAGA	29340
AATTTTtagc ATTTCTGAAT TTTTCTTCAG TAAGTCATTG GTGGTTTCAG AGACTTGACG	29400
TTGAGCAGTC ACAGCATCTT TTTGGCGGAG TAAGGTTAAG GCGATGACTA CTTGGTTTTT	29460
CCATAACGGA ATCGCCGTAT TCACCGATGA TTGAATTTTC TCAGCCAAAG CTTGGTTAGT	29520
ATTTTGgATT AAACGAATTT GTGGGGCTTG TTGAATCGTA ATTTGACGAG CCAAGCGTAA	29580
ATCATGGGTT CTTTTTCTA AACGGTCTAG GAATTGGTTA TAATCATTAA CAATTTGGAC	29640
ATCCATTTGA TCTCCAGATT CTTCCGCTTT TTTCATGGCT TCAGGGATAA TTTGCGTCCG	29700
TAGTTCTTCT GCTTTTAATT CCCCAGCGGC AATATAGATG TTTAAGGCAT CAAAATAATC	29760
TTTATTTTTT TGATACAATT GATCTAACAT GGCATTATCT TTTAAAAGAC CATCTTTTTC	29820
ATGGGTTAAT TTGACAGCAA TTTTGTCAAT TTGCGCACCA ATTTTTTGAT ATTTGGCGGT	29880
AATTTCGTAA ACGGATTGTT TGACTTTACC TAACATTTTT TGAATAAAGT TGCCTTCGCC	29940

AGCACGTAAT TCATTGGGAT TGGCTTCATT TAAACGAAAC ATTAGATCGG TCAGTGAATC	30000
ACCAACAGGA CCAATGTCTT GTGCTTGAAC GTGGTTTAAC ATCGCTTGCG AAAATTCGCC	30060
GAGTTTGATT TGGGCATTG AACCATAGGT AATTACTGCT TGTGAATCTT GGACATCAAT	30120
TTTTGAAGCT AATTCCTTTG CTTGGGCTTG ACGTTCTTGT GGTAATTTAT CAATTAAACG	30180
AGGAGCTGCT TGCTGTTCTT TTAACTCAGA TAGTTCATTT TGTTGCGTTG GTGTTAGTTC	30240
ATCCATTGAT GAAAATGGAT TACTTAATAA ATCGTCCAAT GTATTGGTTA CAGATGTTGT	30300
TTCTTTGGGT TGTTGCTCAG TCATTGTTTT TCCTCCCTAA TGATTATTGC TGGTCTTCGT	30360
TTACTTGAGG TGATTCTCA GTTGCAGCTT TTTGCGACAA GCTGCTTTTA GCGATCGACA	30420
TTTCGACATC TAAATCGTCT AAGTCATCGG AAACGATTTT CTCATAATCA TTTTAACTA	30480
ATTTTGACAA TTGGTCAATG ATTTGTGCGC TTTCTTCTAA TTTTTCATAC GTTTGTTTGT	30540
TTTTTACTTC GTGTTGTTTCG ATTTCTAAAT GTTTACTTGT TAAGTCAACG ATATTTGGTA	30600
AATGTGTATA GAGAAAGTGA TTGGCTAAGT GTAACTTTTT AGGTTCTTTC ACTAACTCTT	30660
TAAACAGAGC TTTAGAAACC TTCGTAGTAT CATTACGTAA GTCAATCGCC CGTAATTTAG	30720
TTGAACGATT CATGTTTTCT TGCAATTGTA TGATTTGTTT TTTGGCTGTG CTCATTGTAC	30780
TTCTGAAAAA ATCAATTTCT TGTGGAGATA ACCCCAACTC ACGATAATGG GCTTCTTTAT	30840
CTTTCGTAA GTAAGGCATT TGATCGCTTT TGTCGGCGGC ATTTTCTTT TTTGATAGA	30900
GAACGAATAC TAAAAAGCCT ACCAGAATCA CAAAAGCGAA GAAGCCATAT TCTCTAGCGA	30960
CAGcAATGaT TAAGGCTCCT AAAAGAATAT AGCCAATCAG TAATTTTGTA TTTAATTTGT	31020
TCATTTATAT AGTTCCTCCT ATCTATCTAA TCCTCGAAAG AGTCAGCTCT CGTTACAATA	31080
GTTATTTTAG CATACTCTGT GAGAGAAAAG ACATACAGCG GAAGATGTAT TTTATGATTT	31140
AGTGAAATCC TAAGAAAAAA AGTATAAAC GCAAAAAAAG ACCCGTCCTA ATTGGGGGTA	31200
GGACGGGAAA GGAGTTAAAA ATGAAAAAGT GTTTTGTTAG GGTGTTTGT TGGTATGCTT	31260
AAATAATACA AGGAAGTTGT GAATTTTTTA TGTCTATTTT TTTTCGAAA TGTAATTTTT	31320
TTCTAAAGCA ATAAATTTTT TTCTTTGaAA AGGGTTTAAC TAGAGAAGAA GGGGTGCGAT	31380
TTTTATCATT TTAGGGTATG ATAAACAAGA AAGGTTGGTG AAAGtATGTT AAATTATGAG	31440
GACTTCGAAG rAAAAACACT TCmAcgTCGA GrAATTTTTTA AAGGrAAGrT TATTGATGTT	31500
TTTTTGATG ACGTGGCTTT ACCGACaCG GCACAGCTAA ACGCGAATTG GTTTTTCATT	31560
CAGGCGCAGT AGCTATGATT CCGTTAACCG CAGAAGGAAA AATCGTCTTA GTTAAACAAT	31620
TTAGAAAACC CTTGGAACAA GTAATTTTAG AAATTCCGGC GGGAAAAATT GATCCTGGAG	31680
AAGAAAATCA ACTGGAAACG ACGGCTATGC GGAATTAGA AGAAGAACT GGGTATCGTG	31740
CAGGTCAGCT TACCTACATA AATTCAATGT ATCTGTCACC AGGATTTGCT AATGAAAAAT	31800
TGGCTCTCTA TTTGGCAACG GATTTGCAGA AAGTCGAAAA TCCGCGTCCT CAAGATGAAG	31860
ATGAAATCTT GGAACCTTAC GAGCTAACAA TAGCTGAAGC CAAAGCAGAG GTGGCAAAAG	31920



337

GCACGATTTG TGATGCCAAA ACACTTTTTG CTATACAATA CTGGAATTG TATTTACTGC	31980
AGAGACAGTT TAAGGAGGAC ACGCAATGAG CAAAGGTCGG TTAGTCACTC GGACAGAGCT	32040
TCGCAAACGC AGAGAAGCAG AAGAAAAAGA AGCGGAACGT CGTCAGCAAG AAGAGCAGAA	32100
GCTGGCGGAA AAAGCGTATA AGCGAAAAGA AAAAGAAATT TCGACGTTTT ATCGTAAAGA	32160
AAAGAAAAAA CaAAAACCGA TCAACAAGTC ACGAGTAGGA GAATACTCGA AGCGTCGAGA	32220
ACGGAGTACT TGGTTAAACA AGGCAATTAT TATTGTAGCG ATTTTATTAG CCGTTGTGGC	32280
ATATATCGTT TTGAATTTAT AGAAAAGAGG AATCACTATG AAAATTGGAA TTATTGGAGC	32340
AATGGATCAA GAGGTCAAAA TTCTAAAAGA AAAATTGACA GACACGATGT CATGGGAACG	32400
AGCAGGCGCT TTATTTGTTT CTGGTTCGTT AGGAAGACAT GAGGTGATTG TGGTTCGTTT	32460
AGGAATTGGT AAAGTGGCCT CAGCTGTGAC CACGACTTTA TTGATTCATC AATATGGCGT	32520
AAATATGGTG ATTAATACAG GTTCTGCTGG TGGTATCGGT GAAGGACTAA AAGTTGGCGA	32580
TTTAGTCATT GCAGACAAAT TAGCTTATTT TGATGTTGAT GTCCTGGTT TTGGTTATGC	32640
CTATGGACAA TTGCCTGGTG GTGTGCCTTT GTATTTTGAA ACAAGTGAAT ACTTGC GGAT	32700
GGAAATTGCG AAAGCCGCTG AAAAACTGG TTTAGCTATT CGTAAAGGAT TGATTGTTAC	32760
TGGAGATA	32768

## (2) INFORMATION FOR SEQ ID NO: 18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6930 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AAGTAGAACG TAAATGGTAC GTAGTAGATG CAACAGATGT TCCTTTAGGA CGTCTATCAG	60
CAGTTGTTGC TTCTGTACTA CGTGGTAAAA ATAAACCAAC ATTCACACCA CATGTGGATA	120
CTGGAGATTT CGTAATCGTA ATTAACGCTG ATAAAGTAAA ATTAACAGGT AAAAAAGCTA	180
CAGACAAAAT TTACTACCGT CACTCAATGT ATCCTGGAGG ATTAAATCA GTGACTGCTG	240
GTGAATTACG CGACAAAAAC TCTCGCCGTC TTATCGAAAC TTCTGTAAAA GGTATGCTAC	300
CTAAAAACAC TTTAGGACGT AAACAATTTA CTAAATTGAA CGTATACGGT GGAGCTGAGC	360
ATCCACATGC TGCACAACAA CCAGAAGTAT TAGATATCAC GAACTTAATT TAAGGAGGGA	420
AATTCATTGG CACAAGTACA ATATAGCGGC ACAGGCCGTC GTAAAAATGC AGTTGCCCGC	480
GTACGCTTAG TACCAGGTAC TGGTAAAATT ACTGTAAACA AAAAAGACGT AGAAGAATAT	540
ATTCCACATG CTGACTTGCG TGAAGTTATC AACCAACCAT TTGGCGTAAC TGAAACAAAA	600
GGCGCTTACG ATGTAATCGT AAACGTAAAC GGTGGTGGCT ATGCTGGACA ATCAGGAGCT	660
ATCCGTCACG GAATCGCTCG TGCATTGTTA CAAGTAGATC CTGACTTCCG TTCTGCTTTA	720
AAACGCGCTG GGTACTTAC TCGTGACGCA CGTATGGTTG AACGTAAAA ACCAGGTCTT	780

AAGAAAGCCC GTAAAGCTTC ACAGTTCTCA AAACGTTAAT ATATCTTATA TATCAACATC	840
TCAAGACACT TTCCAAATTT TGGAGGGTGT CTTTTTTTGT GTTTTTAGGT ACGACAATGG	900
GTACGAGTTT TTTAAACCAG CTATTCATTT TATATTTTTT GCTTCTGCTA ATCGTTGTTT	960
AAATTTTTC TAAGTGGATA GATCAACGAT GTTTGGTGTA TTCACATAGA CTTCTGTTGT	1020
TTTTGTATCA GAATGGGTGA GCGCTCTAGA AATTTGTTCC ATGGTTGCGC CACCTTCATA	1080
TGCTAGCGTA CTAAAAGTAT GTCTTAACTT ATGAGAAGAA GTATTTATTA AGTGTGTTGT	1140
TCGACGTTTA ACTGAGTTGA TGCGGTAGTT AAGATAGTCT ATGTGTACCG GGACATTCAC	1200
ATTTCCCTTT CTATCAATAT AAGTGAAAAG AAATTGCTCT GCATTTTGnT tAATTCCAAG	1260
TTGTAAAAGT TCTTGTGCTT GTTCTTTTTT CCATTTTCATT AACAACTCTA TAAGGGAGGA	1320
AGGCAATTGG aATTTAGTGT TTTTCTTCC TTTGGTAGGA GTTAAATTC CTAGCTTGwC	1380
TTTGCTTGG ACAAGTAGTA TTGTTCCATT CTGTAAATCG ACATGTTTCC ATTGCAGAGC	1440
ATAACTTTC CTTTTACGAT CGCCTAAAT TAATGTTAAG GTAAAAAGGA CAAAATCTTG	1500
AAAGGTTAGA TGTTTGTTAT TGTAGTCTTG TTGAATGGCT TGAATCCACT CAAGAAGTTG	1560
TTCTGCGGTT AGTGCTTCTC CTTTTAGCGT TCACTCCATT TTTCTACGTT GTTTTAAAGG	1620
ATCTCTACG TAACGAGTAA TCTTAGCAAC ATGATTGTAT TCGATGTAAT CAAGAAGTTC	1680
TGCAATATCA AACAAATGAT TTACATAGCT TTTAATCGTT TTGATATTG CATATTTTTG	1740
CGCTTTCGCA TTTAATTTTC GCAAACTAA ATCTTTATTT GTATTCAATT CAAAGAGAGA	1800
GTAACACCG AACATAGGTA GAATATGAAG ACGAAACAGG TCTTTAGTGT TAAAATCGT	1860
TTGTTCTGAT GGAATCATTC GGTTCCTCCC ACTAGAACCA TTGCAATACA TGTCTAACCA	1920
GACCTGTTCA TAAAATCTT TGAAAGAAAT ATTTGCTTTT AATTCAAAAG CTCGTTTCATG	1980
TTTTTCAAGT TGTACTTTTT CAATTTTTTT AAGAATATCT TTTTCAGCTA CCAGAGCTTC	2040
ATTTTGGAA GAGAACGTTT TTCTGTATCG CTGTGAAGTA ACACCCAGAA TTTCTCTTAC	2100
TTCTTTAGGA TAAAATACTT CGACTTTATA ACTTCCGCTT TTCATTTTTT TGATTGCCAT	2160
TAATATTTCC TACTTTCTTG CGTGAATCAG AACGTAGAAA GACTATCGTT TGATTATAGA	2220
ATAGATACTT GAATGTTTAT AGGTTGGTTT AAGCATATTT TTTTAACCAA TTATCAATCT	2280
CCAACCTATC ATAAAGTACT ATTCCGTAA TACTGATTCG TGGAAGTCCA ACTGTAATTA	2340
ATTTATCAAT AGTATTGTTG GAGAGATTAA GATATTGGCA TAGCTGATGC TTTTAAAGAT	2400
AACGATTATT TAATCTTAAA TCAATTTGAA ATTGTGTGAT TTCATTTTCT AGTAGTTTGT	2460
GAATGGCTTC CTGTAATCTT TGTTTTTGAG TGTGAGACAA TAAAATAGTG AGTTCGTTCA	2520
TTTATGCGTT CCTCCATTTT TCTTGATTTT TGTGTAACGA ACATTTTGTT CTTCAAAC	2580
TATTTCTTTG TTTGGGATAT TTTGTTTCATT CCAGTCTTGA AAGAAGACAG GTCTTTTACT	2640
TACATAATTG TTGGTAAATA AGACTTTTTG ATCTTTTACT TCATAAGGTC GATCTGTCAG	2700
TCTTTTATT TCGAGCGGAT GCTTCAAGTT TTGAGATTTA AAGAAGGCTT TTTTCTTATG	2760

CTCTTTTAGT	TCTAAATCTT	TATCAAAATA	CTTGCTGACA	TACCGTCCTC	TATTTTCAGC	2820
GCTATCTATG	TCAATTTTGT	TGATTTTGAT	TAGTCCATGT	CCCCAAATAG	CTGTTAGTTC	2880
TGCATGAGGT	AAATAGGGGA	AGCTGAATAA	AATAACGTGA	TAATGAATAG	CGCCATGTTT	2940
TTGTTTCTCC	CACGTAGCTA	GGTACTTAAT	TTGGGCTTTC	TTTGTCTTGT	ATAAATGGTA	3000
GTTTAGTCGG	TTCATAAATT	TTTGAAGTCT	ACTATTTGTA	TAGGCTATAT	CTGTTATATT	3060
CTTTTTAAAA	GTTAAAGTTA	AAAATTTTGT	CTGGTTATCC	AAATTCATAT	CAATGAGTCG	3120
AGCAACGACA	AAGCGTTGAT	TTTTGTAATG	TTTTTGTTTT	CTCTTTAAAC	TGTCATATTG	3180
TCCTTGTCG	GTTAAATCAT	CAAATGTCTG	TTTTGTAGTT	TTTAGCCAAT	CAGGTTTTTC	3240
TTCTTTTTCA	TCATTATCTG	TTTTTCGTTT	TGTTAAGATT	AATTGTTTAT	ATTCCCAAAT	3300
TTCAATGTAC	GTAGGGGTTT	CAATGATTTT	TTGATTGTAT	CCATGTAGCA	ATCGCAATCA	3360
CTCCTACTAC	AGAAATTGAG	TGTTCTTACT	GGTTATTCAG	TTGTCAAAGA	ACGAGTGTAT	3420
AACTTTTTGA	TTTTTCGCCA	GATAATGTTC	TATAAATCAA	GTTAAGTAGG	AAGAAAGCCT	3480
ACGCCTTTCT	TCCAGAGCAA	TCATCCATCG	TGTGGTTGAT	TTTCGAGCGG	AAAGCGGCTT	3540
GAGCTTTGAT	GGCTCTACCG	CTTTCGTCGA	AACCACCCGC	TGGAAGTCTG	CTCTTGCGTG	3600
GATAGTGATT	CTTCAAGTTG	AGCCAAGCGT	TGTTCTAATG	TTTGATAGCT	TCTTTCCAAT	3660
TGTTGGTACT	TTTTTGTTAA	TGCTTGAAAC	AACTCCTCTT	TATTTGCCAA	CTTAACTTCA	3720
ATTTGATTGA	TGATTTTCACT	TAGTAATTCA	CTATGATTCA	TGGTATCTAA	ATGATGTTTC	3780
TTATCAAAAA	GTTTTCGTAA	AGAAGTATTT	CTGCCCAATA	TAGGGATCAG	TTTTATTACT	3840
TTTTGAAATT	TTTGGATTTT	TTCCAGTGAA	AACATTCTGG	TTTTGCGGTG	ATTGATCGTT	3900
CTTTTAGTTC	TTGTGTGACT	GGCAAAACGA	ATATAATCAT	AGTGAAATTG	GTAATCGCAC	3960
TCTTTTTGAA	TAAAATTTGC	CCATTTGTTT	AAAGTAGAGA	GGGAATTCAG	CCCCTCTACT	4020
TTTTTTAGCA	CGTCATGGGC	GGTGAAGTAA	TCAGTCATTT	GTTTGGAAC	AGTGTGGCTT	4080
TATAGGCTGT	AGCAGTTAGT	CCATAGCTAA	CAAATTGTCC	ATTTACACGA	CCCCAAGGAG	4140
TAATCATAAG	ATTATCAAAT	CGAACGATTG	CCATACCTTT	TTCATCTAAT	GTTTCTTTTCG	4200
TGATTACAGG	AGCTTCACTT	TTAACTTTGA	CACTGATTTT	TTCAAATCGG	AGTTTTGGAA	4260
GAACACATTC	ATACGTCCAG	CCCAACGTTT	GCCGTCTTTT	TGCCAAAGGG	AGACTTGGAC	4320
TACTTGGTAT	TCATTTTCAA	GCTTTTCTTT	AGGTAAAGTT	AAATCGTTAA	TTTTTACTGC	4380
CATTTTCATAT	CCTCACTTTC	TTTTAGTACA	TTTTTTAGTA	CAAACATATT	CTAATGGAGC	4440
TAGAAAAATT	TGTTAATATT	TTTTAGCACA	AAAATTAGTA	CAAAGCTGTT	TTACTCTTAG	4500
TGCATTATTT	GTTATGATTG	TTTTGATAAA	AATTGGCTTT	TAACCATTGA	GGAGATTGAG	4560
CTTTATGAAA	CCAAACAAAA	AAAAGTAGGC	GAAAGAATTC	GAGAAATTAG	GACAAATCTA	4620
GGCTACAGTA	TGGACGAATT	TGGAAGTTTA	TTAGGTGATT	CTCCACGAAT	TCTGTGAACA	4680
ATTGGGAAAA	AGGGGTCAGT	ATTCCTAAAA	GAGATAAACT	AGAAAAAATT	GCGATATTAG	4740

GGAATATGTT	ACCAGACCAG	ATACTATATG	GTCGAGCAGA	TGAATATTTA	TATGATTTAA	4800
TTGATAGCAG	TTTCAATGTT	AAATTCAGCG	ATAATATTCT	TTATGAAATT	TTTGAAAGTG	4860
TACCACCTGA	AAAGCGTTCT	TATGATGATT	TGATGTGGTT	AGCCGTTGCG	AAATATTTTA	4920
TTGATAATGG	TACATATGGA	AAAGTTGTTG	GAAAATTCAA	TTATATTTCA	ATGCTAGGTA	4980
TACCGAACCT	TTATACAGGA	AGCTATCAAA	ATGAATTTAT	TGAACAACGT	GAAGATTATA	5040
AACAACAAGA	AGTTAAATAC	TATGTATATG	CAGATCCGGA	AAAAACATA	TTACACATTA	5100
TGCCATTTGT	TCCTAATGAG	AAGAATATAG	AAC TGCTTTA	TGAATTTCCCT	GAATTATTAG	5160
AACAACAAGC	AGATCACGAA	GTATTTACTG	CAAATTTTGA	ACAAATTGGG	CTTACCTTAA	5220
AAGGATCAAC	CATTGTGTAT	TATGGAATTG	ATAAACAAAA	ATTAGAAGCT	AAAACACAGG	5280
CATATAAATA	TGATGAAGCG	GTTGATTTAT	ATAGAATAGC	TCAACTACCA	GAACTTTTAT	5340
CAATGTTTGA	TGATGAAATC	CAAAAGCAAA	TTTCGTTTTT	AAAAAGTAAA	TAATTGCATA	5400
TAAAAAGAAG	TCAACAGTGT	TTCATAACTG	TTGGCTTCTT	TGTTTTGCTA	TTTAGTTTAT	5460
AAGTTAGGCC	AGAATCAAGT	ATATCGTAAG	TCTAAATCT	AACCTTTTCT	CTTCTTTTAA	5520
AATTTTCAAG	CTGTTTTTCT	AACTCTCAG	TACGTTGTAG	AAAGTCATTG	AGAAAGATTT	5580
TGAATATTTG	GGCATTTGCA	TGGTATTTTC	GAATGATCTT	TTTTTTATTA	ATTGTTTCTT	5640
GTTCAAAATC	AAAAAATTTG	TTTAATATCT	CTAATGTTGA	TGTCCATGAT	TGAAGTTGTT	5700
CAAGTGTTAT	TTCTAGCGCA	TTGATATCTT	GTTGAGTAAT	GATTTTAAATC	GGCTGATTCT	5760
TCTTTGTTAG	TAATCTTTTA	TTTTCTTTTT	CCGATGTCAT	AAGGTTTCCT	CCTGCTATTT	5820
TACAAAATCC	ATGTCTATAC	GAACAACAAT	CTAAATTAT	TATACTATAC	CATGTTTATA	5880
TAGACAACAC	GGGGAGTGAA	TATGAATGAA	CGAAATGAA	GGTGGTAGAA	AAGTGGAAGT	5940
GCTGATCAAC	GATTTATTAG	CAGAACGTCA	AATGTCTTTA	AGAGAACTCG	CAAGATTGTC	6000
GGGAATTGAG	CCTTCAAATT	TAAGCAATT	GGCAAATGGa	AAAAGACAGA	AAATTTATTT	6060
AGAACACATT	GAAAGAATCG	CAGATGCGTT	GGAGATTGAT	GATATATCGA	GGATATTAAA	6120
GTTGAATAGG	AAGTAGTATT	GGGGATATAA	AGAATAAATA	TAAAAATAGT	TAAAGATATT	6180
GTTAAATACT	TTAGATTAAA	TGAAAAAATT	ACAGAAATAA	ATTTGATAGA	AGAAAAGTAT	6240
AAAGCGAACT	ATTTGATTTT	TAATATTTAC	TGATGTTTAC	GATGAAGGTG	GAATATAAAT	6300
GAAGAATGAT	AACACTCTTT	CTGAAGAAGG	AATTGATACA	GATAAAGGAA	TAGTTGGTTC	6360
AATTTTTGAA	AGTATGCTAT	CGGATACTTT	ATCAGAAATT	AAAAATTTAG	ATAATGTTAA	6420
AATACAGAAG	ATAGTAGAAA	ATATAGGGAA	ATATGATAAA	AAAGAGTTTT	TGCAAAGACT	6480
TGCTGCACTA	AGAATTCCTT	TTGAAAATAG	AGATAAGGCT	GTATTATTGG	ATGCAACAAC	6540
TACAGCAACA	CTGAATTGGT	TATCGGAAAA	CAACTGGAAT	TTTAATGGGC	TTTCAATGAG	6600
CTATGGTAAA	TTCAAAAAAG	TTATTCAACA	GATTAAACCA	TTAGACTCTA	AAATGGCTAT	6660
TGATCCATTG	GATAATCCCT	ATATAGATAA	TATCCaaTTT	TATGGTAATC	ATAAGGTAAT	6720

GCCTGGAATA AATTTTGCAA GCTCGTATAA TTTGCAAATG ATGATCCAGT CTATATTTTT	6780
ATCAAATCGT ACAAAGTTAT CTGATGAGAA AAATAGATAT ATATCCATTT CATTAAATGA	6840
TAACCTTATG GTTTCTACCG ATGTTTGTA AAATGTCAGA TTCCAAACGA ATTnCCATCA	6900
TTCACCTCGCG AAATATnTAT nCCTAATAAA	6930

## (2) INFORMATION FOR SEQ ID NO: 19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25580 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TGTACTTTCG TCCAATATTT ATTTCTTCTT CAAGCTGATA AATAGACACA TCTCTATCTT	60
TTGCTAGTTT TTTTATACGT TCCAAAAGGT TCATATGAAA ACATTCTCCT TCTAGTTACg	120
mGTTATTTmC AAAAAAAGTT AGTAAAAATG TTTGACATTT ACAATATTTG TTAGTAATGT	180
AATCACGTAA GCTAATTTAT TAGCTAATAA GTTCTCAAAT AAAACCTATA AACAAAACTT	240
GAAAATCGTT GGGGAACGGT AAAAGTATTG TTTTAGAAGG CTTTAAAGT CTTATTTAGC	300
TATGGGTTCA TTTTACAATA TGTGTTAGAA AGTGCAACG AATTTTATTA AATTAGCTAA	360
TTTTTTAGCT TACAAATTAA AAAGAAGGAA TGGAAGAAAT GAAAATTGAC GAaGCTGTTT	420
CAAAAGCAAT GAAAAAAGGa AAATATATTT ACAGGGAGTC TGAAAATGAC TGTCAGCTC	480
ATGTAAATAT ACTTCCTACA AATACCTATG ATTGTTGTCT ATTACTACAA GAAAATAGTA	540
ATTCTGTTGG CAAACGGTGG AACCTACAG CAAATGATTT AATGGCGAGT GACTGGGAAA	600
TACGTTAGTC AAGGCCAGT GTTTTTGTAA TAACTTTAGC TGCGATTTCA CTCATTATGT	660
CAATTGAGAC ACTTGcTAGT TTGGATGATA TnTTTTTTGA TTCCTTCCAA tTTTCGGATC	720
TCGAATATTG TCCAAATATT TATGACCATC GAATGTAAGG GAAGAGATAC ATGTCATATA	780
AGGTTTGTTA GATGCAAACT TAACATATCC TTTTATAAGT TGTGCATCGT CATCACCTAA	840
TTTGAGACG GCATAATTAA TTCCTCTGA ACTATATTTT TTTAGTAGAT CTGATTTAAA	900
TATTTTATCA CCAAAGCCG GTTCATTGTA CGGAAGGCTT TCGCTAAATA ATAAAATATC	960
ACGAATACAA TCATGAGACA GTTTCATATC ATTCACCACC TTATCAGTTA TTCAGCAGA	1020
CCACTTGCTG ATAAGGAAAT TATATCAAAA AAAGAAAGTG AGGTAATTAA ATGTCACAAG	1080
ATTTAGCTAT TGAAGTAAGG GCAGCGCTAA TTCGTGCAAA GAAACTCAA TCTTGGTTAG	1140
CGAAACAATT AGGGATTTCA AGTCCGTA TATCTGATAT TCTTCATGGT CGTAGACGTT	1200
CAGAAGAGCA AGTTCGGAAT ATCmAAAAA TATTAGATAT TAGATAGGAG GTGTAGGTAA	1260
AAGTGAAGA AGCGATTATC AAAGTTGACT TACAAAACTT AAAGAAGTTG ATCAAACAAG	1320
CAAAGGAGCA AGCTGACCAA CTTCAAAAAA CTTTAGATGA GATAAATAAA ACTAAAATCC	1380

TAATTTCTTA	GCTACCCATT	CTGATCCAGC	AGCAGACTTC	ATATCTTCCC	ACGAATCGAA	1440
GTTTGTGTTT	GAGCTAATGA	AATTATCTAA	TTTATTGTCA	TCTATTGATT	CCATATCAGA	1500
GAAATCAAAT	CCGGATTTTT	CAATAAACTC	ATCAATATTA	GAGAATTTAG	TATTTTCAAT	1560
CATAAATTTT	TTAGTGAATA	GTTTATCGAA	TGGTACAGAA	TGTTTCGTCAT	CTAAAGACTC	1620
GCCATTTTTA	GCAAATTGAT	TTAGCTCATG	TTGTAATTCG	TCAAATCCAT	TTAATTCGAA	1680
TTTCATAATA	TTTCACCTCG	CTTTCAAATT	CATTTTACCA	AGAGGTGAAT	TGTAAAACAA	1740
TCAGTATAGG	AGGTGATAGC	AATGGAAGTG	ATTTTAACTC	CAGAAAATGA	AGCTTCTCTA	1800
AGAGATTTTG	TACACGGAAT	TATTGTTGAT	GAAATAGAAA	AAGCaCGaAG	aGaTACCGcA	1860
GTTGaTAAgC	gaGtCtTAAA	tCaAACaGaG	rTTGcaAAAT	AtTTcaATGt	TTCCACAACA	1920
ACAATAAGGG	AATGGGAGAA	GCTAGGTCTT	CCACATGGAT	CAGTAAGTAA	ACAAGGGAAG	1980
TTCTACGACA	AAGAAGGGTG	TCGGAGATGG	CTTCTATCAC	AAAAAAGATA	AATCTTGGGC	2040
AAGCGAAATC	AAGGGAGGAA	ATAAATATGA	AAAAAATATA	TCACTTAAGA	CGTATGGCGG	2100
CACTGTTGGT	TGTATTCCGA	CTAGGTCTAT	TGGTAGGTGG	CAATATTGGC	CCATTAATCC	2160
AAAACATATA	TATAGCAGCT	TTTATCATTT	GGTTGCTCTA	CTACGATTTA	GCGTTGGAAG	2220
ATCGAGAAGA	AAAAAAACAA	AAATAAGAC	CCACTTCGAC	GGCCATCAAA	GTAGGTCAGT	2280
TACAAATATC	AAATTCAAGG	AGAGTGTAAC	ACATGAATAG	AAAAATTGAA	AGAATGATTA	2340
TTGAACTTGA	AAAAGAATGT	AAGGCACAGA	ATGTTGAACT	TCTTTTATGT	GCTACAAATT	2400
TTGAAACAGG	CCAAGGAAGT	ACTGCGTTTT	GTGGTTCAGT	TATCGGTTTA	GCTATACTCT	2460
TGCAAAAAC	TGTAGGTGAT	CTAAAAGAGC	AATTAAGTAT	AAGCGAATCT	TGTGATTGTC	2520
CAGAATGTGT	AGCAGAAAAA	GCCGAAGATG	CTGCAAATGA	AAAATCTATG	GATGAACTAC	2580
TAAGTGCATT	TTTACGAGGT	GAAGTCAAT	GATTGAAGTA	AGAGGTTTAA	GTGATGATGT	2640
TTACGAATTA	ATGTTAGCGA	ATGCTCAAAA	TAGGATTATT	CAATCAATTC	GAAGTGCAGC	2700
AGCAAATGGT	AATACAAGTT	GCGTAGTGAA	TAGTAAAGGT	CTTACATCAA	CGTTTTTATC	2760
TCAATTAGAA	ACAGAAGGAT	TTGATCACGT	TGAACCTGAA	GAAAACAAAA	CGAAAATATT	2820
TTGGGAGTGG	TGAAAATGCC	TGAATTTGAT	TCATTAGGGG	CTAGACAAGA	ACCTCCAGAA	2880
GAAAAAGAAG	CATTAGAGCC	AACATGGGAA	TATGACGAAG	AAGAGGAGAA	TGACAATGAG	2940
TAACGATTTA	ACACAAATAA	CACAACGATC	TTTAGATGAA	CAAGTCATCG	GAAATTTGAA	3000
TAGATTGCAA	GAGCAGGGAT	TAGAAATGCC	ACCAGGTTAT	AGTCCACAGA	ATGCATTGAA	3060
AAGTGCTTTC	TTTGAACTAA	CCAACAATTC	AGGAGGAAAC	CTTCTTCAGT	TGGCAGCTAA	3120
CAATCCAGAA	ACTAAAACAT	CTATTTCTAA	CGCCTTGCTT	GATATGGTCA	TCCAAGGATT	3180
ATCGCCAGCT	AAAAACAAT	GTTATTTTCAT	CAAATATGGA	AATAAAGTTC	AGCTTATGCG	3240
CTCATATTTT	GGAACCATGG	CTGTATTAGA	TCGAGTAACT	GGAGGGGCAG	ATATCACGCC	3300
TGTTGTAGTA	AGAGAAGGCG	ATGTATTTGA	AATTGCTATG	GATGGTCCCG	ACTTAGTTGT	3360

TGCTAAACAT GAAACGGCCT TCGAAAACCT AGACAACGAC ATTAAGGCTG CTTATGTGGT	3420
TATTAAGCTA GCAAATGGTA AAGAAGTAAC AACCGTCATG ACAAAGAAAC AAATTGATAA	3480
GTCATGGAGC AAAGCAAAAA CAAAAAATGT TCAGAATGAT TTTCCAGAAG AAATGGCAAA	3540
AAGAACTGTC ATCAATCGAG CTGCTAAATA TTTAATCAAT ACTAGTAACG ATAATGATTT	3600
ATTTGTGCAA GCCGCTAAAG ACACACTCGA AAATGAATTC GAACGAAAAG ATGTAACACC	3660
AGAGCGAGAA GAGCAAGCTG CGGTACTTGA AGAAAACTA TTTTCCAACA ATATAAAGC	3720
TGTTGATCAA GAAAACGAGA ATGAACGAAT TACACGTGTA GCTGACGTAC CAGAGCAACC	3780
CGATATTGAA CAAGCCAAAC CAATTGAAAA AGATAATTTA ACGAAAGTAG CTGACCAAAT	3840
TTTAGAAGAA CCAGTTCAG AACTTTAGA TGTGATGGCT GGTATGAAA CCAATCAGAA	3900
AGAGAGTGAA GCTGATGTCT CAACGATTGA AGAAGACGAT TATCCTTTCT GATGAAAATT	3960
ATTATTCACA AGAAGCGGAC CTAGCTTATA TGTCTGTCTC TCAATATAAA AAATTTCTTG	4020
AATGTGAAGC TGCAGCTCTT GCCAAGTTAA AAGGCGAATG GACACCAGAA AACGATCCTA	4080
AAGCATTGCT AGTTGGTAAT TATGTTCAAT CTTACTTTGA ATCACCAGAA ATTCATGAAG	4140
CATTTAAAGA AGAAAATAAA AGCAAGATGT TTTCTTCAAG AAAACCGTTT GGCCTACTGA	4200
AAGATTTCCA AATTGCGGAG CAGATGATTG AAAGATTAAA ACAAGAAGAA GCCTTTTTAA	4260
ATATTTATCA AGGTGAAAAA GAAGTGATCG TCACAGGTGA AATTGGTGGT GCAATGTGGA	4320
AAGGGAAAAT TGATTGTTG AATTTAGAAG AAAAGTATTT TGTAGACATC AAAACAACCA	4380
AAGATATGCA TGAGAAGAAA TGGGATGAAC GTTTAAACAG AAAAGCAAAC TTCATTGAAC	4440
GCTTCGGTTA CGTGTTACAA ATGGCTGTTT ATTGCGAACT GCTTCGGCAA CAATATGACA	4500
AAATTTTCT TCCTCTCATT GCAGCCGTTT CGAAACAAAC ACCTAGTGAA GCAAAACTAA	4560
TCCTCTTAG CGAAGAAAAA ATGATTTACG AATTAGAAGA ATTAAAAGAA AACATCGAGC	4620
ATGTTGTGCG AGTGAAAAAC GGCGAAGAAG AACCAGTTAG TTGTGGGATT TGTGAATATT	4680
GTAGAGGACA CAACAAAATT ACAAATTTTA CCAGTATGGA CGATTTATAA AAGGAGGCGA	4740
GCAGTTGGC CATTGGAGGT TGGATAAAAC TTTATAGGAC CATTCAAGAT CACTGGATTT	4800
GGGAAAATCC TCAATACCTT AAATGGTGGC TTGATTTAAT ATTCATGGCC AtCACCAGAC	4860
AGAaAGATTC TTTTGTATGG AGAGTTAAAA cGGTGAAAGT CGGAGAAAGA wTTACATCCG	4920
AAAAAAaCT TtCTGAAaGG TGGGAaGTAG CCGAAACACC GTCAGAAAAt TTTTAGATTT	4980
GCTTGTTAAA GATGACATGA TTGAATTAAG CAGAAGTAGA CAAAACGGGA CATGGTACAA	5040
AGTCAGTAAC TACGCAGAAT ATCAAACTT TTCTGAAATA AAAAAACAAC GAACTGAACA	5100
ACGAAGGGAA CAACAAGCTG CACATCAAAC TGAACAACGA ACTGAACATA AACAAGAACC	5160
TAAAGAACCT AAGAACCTAA GAAATAATAA TAATAACAAT AAAGGGGCGT TCATTTCGTT	5220
AATTTGGGAA AATAACGGAT TTGGATTGAT GTCGTCTAAA ACCATGACCG ATTTTGATTA	5280
TTGGATTCTT GATTTTGAAA AAATCGGAGC TAGTCAAAAA GAAGCTGAAC AATTAATTGT	5340

TAAAGCTATT GAAATTGCTA TTGATGCAAA CGCAAGAAAC TATAACTATA TCAATGCCAT	5400
ATTGAAAGAT TGGGAACAAA GAGGGTTCAA ATCTGTTGAG GAACGAGAAG CGGCAAGGAA	5460
GCAAAAGAAA ACAACCAAAA AACAGAAATC AAACACAGGT CATTCTGATT ACGATGATCT	5520
TGGATTTTAG GAAGTGAAAG AATGCAGTCA GCATCAGATG GATTTTCAAA AATGATTAAA	5580
ACGTTGCTTT ATATCACGCC TGATCCATGT CCAGAGTGCG GAGGAAATCT TTATGCTTGG	5640
CGTGCAAAAA ACAAGATGG GTCCGATAGG TGTCCGCCAA CTTGCATGGA ATGTGGCTAT	5700
AAAGCACGCA AAAAGCAGA AGACCTTGAA ACAGAGAAAA TGTTTAACGA TAGTTTGAAA	5760
GCCAGAGCGA TTAATTATCT GAAATATAGT TCGCTTTACA CCGACAAAAA TTTAATTAAT	5820
TGTCGTTTTA AACTTACAA AACAGTAGAC ACAGAACTA AGCTTGCTTT TGAAATTGCA	5880
AATCGAGCCA CAACTGAAAT TCTTTTGAAT AAACCAATTC ATATGATTCT TTCAGGCAAA	5940
AGTGGTGTTG GAAAAAGTCA TTTAGCTATG TCAACGGCTT GGAAGTGTT GGAGAAATCA	6000
AACTATGATA AACGCTGCTT ATTTATTAGC TATGCGGAAC TCTTAGAACA ACTAAAATTT	6060
GCGATGAATG ATGAGCAAGC CAGAAAAGAG ATAACAGGTA GTTTGATGGC AGAAATCAAA	6120
AGCGCAGATT TAGTAGTTCT GGACGATTTA GGAGCCGAGT TGGGAGTTAA ACAAATTGAT	6180
GATAGGAATA AAAGTACTAA CTTCAATAAC GACACCTTGA ATCGCATTGT AGAAGCTCGG	6240
CAGAATAAAG CAACTATTTT TACTACGAAC TTAAGTGGTA AAGAAATGAG TCAAGCTTAT	6300
GGGGAGAGAA TTCTTTCTCG TATCATGAGT AATTCACAAG GTTTTGTGAT GAAAATTGAA	6360
GGGACATCAG ACAACGAGT AGCAGGCATC TAAAATGTTA TTTTLAGCGA ATATATTGAG	6420
CGTAGAGCGG TTTTACAATC AAGTGAATAT AAATAGATAC AAAGAAAGAA AAACGGATTA	6480
AAACGCATTT TAAAGCCTTA AAAACAAATC GATAGAAAGG GGAATCATTC AATGCCGTAT	6540
GTAGTGAAAA TTTCAGCCTA TCTTGGCAAA GATGGTCGAC CTGTAGCCAA TTAAAAGAT	6600
GCTGTGCTAT TTGAGCAAAA AGAGACAGCA GCTATTGCAA CAATCGTATC TGGCGGAACC	6660
GTTTCAGAAG TAAAGGAAGC CATTATAATA CCAGAAAAAC CAAAGAAACA TATAGGGAAG	6720
TCTATTAAAC GGATTGATAA GAAGGAACCG ACCGAAAAG CTACCAAAAG TAATCAAGCC	6780
TGGATGAAAG GGGCTAAATA AGAATGAAGT GTGTTAGATG TCAAGATCAG CGCGTGATTT	6840
GGGGCAAAGA CAGATTTAAT TATGCAACAC CTATTCCATG TCCTGAATGC AACAAAGATG	6900
GAAAAGCAGT TCGAGCGGAA ACTGCGACCA AGGAAAGGGA GTTAAAACAA TGCAATCACC	6960
AACAGCCCTG AATAAGCGAG GAAACAAAGT CACAATTGAT GGTACACAT TTGATAGCCA	7020
GAAGGAAGCT AACTTTTATA CAAAGTTTGT CAAAAATTGT GGGTTACCTT TTGAAGTTCA	7080
TCCGCGTTTT AGACTAACCG AACTTACACC AACTGCGGAT GGTATAGGCA AAATTTCCGC	7140
GATAGCTTAT TCACCTGACT TCATCATAAA AACTTAGAT GGGAGTTGGA GACATGTCAT	7200
TGATATTAAA AACTCTTTTG GCGTGTATGG TATTGACCAA TCCGTTAAGC TTCGTTTTCG	7260
TCTATTTGCC CTTGCATATG GTCATCCAGT TGAAGCGATT GTTGTTCTGT CTAGAGATTT	7320



TAAAGTGATC ACTCAAGGTG TAACTAAGCC TTTAAACGAA AAAAGACCAT TCATAACCGA	7380
TAATTTTCGAT TACGAATGGA AAGATGCAAC TAATTATTAA ACGAAAGTAG GAAAATAAAA	7440
TGACAAAACA AGTAAATTTT AGACCAGAAG TGAAAAAAGT GACATCTAAA TCAAACGGAA	7500
ATATCGAAGT GCTATTAGTG GTTAGCAACG CTTCATTAATA AGGGAAATAT GAAAGTTTAA	7560
ACGAATTTTT AGGTAAAACA GTATCAACGA CCATTGAGCC AGAAACAGTA GAATACAaGG	7620
TACCAGTTAA CAAGCrGACC AwTAAACCGn ATGTCGAATA TATTGyGmnT AACGACGGAA	7680
CAGTTGAAGT TCTAAAAGAA GAACAAACTT CTTTAGAAAT GGGCGATGAT GTGCAAGAAG	7740
TTGAAGAAGT TGCTGTGCAA GTATCGAAAG AAACCATTGA CGAATTCATC AAGAAGGCAA	7800
CAACAATCGA ATGGCCAGAA TCAGTAACAA TCAACGTTCTG TGGCGTGTTG CATCGAATCG	7860
ATGAAGGGGA AGCCTTAGAA GAAATTGCAG CTGATCATGA TGTTCAGTT GAAAATCTAA	7920
TCAACCAAGT TGAACCTGCA CGCCAACATT TTGCACCGTT TGCAGATTCT TGGAGCAAAA	7980
ATAAGAGAA CATCATTTTC CCTGAAAAGA CAGTTGAAGA TGATGAAGAA GAAATCGAAT	8040
AATAATCTCG TAGAAAGTGA GTGTTCAATT TGCTTGAGAT TTATTATACG CCAACATCCG	8100
CAATTATTGC GGATGCACTG GCTAAACAT ATGAAGTCGT TTCTTTAGAA ACAGCTAGAA	8160
ATATTGCCAA GAAATTTAAG GCTAGCTTGA AGCAGAAAAC GGATCTTTAT GTAATTGAGG	8220
GAATTTTGAT TGATGCTGGT TATAAAAAAG AGCCAGTGAA TTTATAGAAG GGAGTGGAGG	8280
TTTGCGGCCG CGATTAAAA GCTTTTTACT CCGTTGAGAT TATGAATATT TACCAAAAGA	8340
AAATGGGTAA AGTTATGAAA GTCCTTTGCG GTAAAAGAAG TGAAAACATT GGGTAAATAA	8400
AAAGTGAAGA AGAGTAGCTT TTTACTACTC TTCTTCGCTA GAAAAAATTT TTGAACAGAA	8460
TAAAAAGCCA AGGTCAGTTA CATGTATGGA GCCACGAATA CTTATAATGC TAAACCCTGT	8520
ATCTTTAAAA AAAGAACTGT TTGCAAAAAA ACTATTAAGA TTATATTTCA AAGTTTTAAA	8580
ATAAGTTGAA TTTTCAAAAT CATCATAAAT TTTTGTATGT TTTTTTGTG TTAATCAAT	8640
AGAAGGTCTT ATATCAATTA ATCCATAAGA GCGTAAAGTT TGTAAAGAGG GACTTAAATC	8700
AGAAAAATTA TTTTCAACAA TTCCTTTCTG AGTAAGGATT AAATTTTCTA TATAATCTAT	8760
ACCGTTACCA TTAGGAGCCT GTACTCTAAC CTTTATAGTT GGGGTAATAC TAAGTTTATA	8820
CAATTTTTTT AATAGATTTG CATCTATAGT CGTTAGTTCA GATAAAATAA TTGAAACCT	8880
TGGCGAAGAA TCTTGATTTT TTCTGTTATC AACTAATCCA GCAaGTAAAT TAGAAAAGTA	8940
TTCACGCATA TCTTCATCTT CTAACCTGATA TCTAGCGTCT TCCATTGTTT TTAAAGCAAA	9000
CCCTAATTTT GAATCATCTC TGTATCTAC AGGGATACTA TCGGTTTTAG AATTAATCTT	9060
TTCTACAAAA TCAATATAAG ATTTTCACT AACAACATTT AGTTTCTAA GAGGACCTAC	9120
AACATAATTT ACAATCCCGC CAAGGCCTTC TCCAATTGAT TTTGTAGCAG GTCCAAAGAG	9180
ATTAGTTATC AATTCCTTTG GTAATAGTTC AATGTTAATT AAAGATTTTT TATCGTCAGA	9240
CATTTAAACA TCTCCTTTTT CAAAATATTA TACCAATAAA GAAAGGAATT TTCTATGCAA	9300

CTTAGACCAT ATCAAGAAGA ATCTCGTTCC GCTGTTCAAA ATGAATGGGA GAATAAGAAA	9360
AAGAAAACAT CATTAGTTTT GCCTGCAGGA TGAATCTGAA TGAAATATA TTGATATGAA	9420
AGTTCAAAAA ATACTGGGTG TTGGACAGAA GGCCTAGATC ACTCAAATAA AAAATTTGAA	9480
AAGGAAAAAG GCTACTTGAT GGTGACGAT CCTGAGGAAG ATGAAACAGT AAAACTAGTC	9540
ACAATTGAGG TGATAAACGT TTTAAAATCC TTAAGAATGA CCGTAACTAT CAAAAAGGCG	9600
ATATCTTACG CTTAAACGAA TATCAAGACG GACAGTATAC AGGTGATGTT CATGTCTCAG	9660
AAATAACATA CATTACAGAT TATACTCAAC AAGATGGTTA TCTAGTGTTA GGAATTAAGT	9720
GAGGAGGATT AATAAATGGA ACAATTCTTA TTAACAAAA CTGGTAAAAA ACAAATTGAT	9780
ATAAAGGGTA CGGGAATGGA TCATAACGAA ATTGTCTTTA CGTTAGCTGC TACTTTAGTT	9840
GGATATAGCA AAGAATTAGG CCTAACAAAA GCAATACTAA ACGAAAGTAT GTACGTGCTG	9900
TGGAAAGATG GTGAATAAAT GAAACGTAAC TGTAAGAGTG TAATAAATAA AGTTAGTGGC	9960
ATTGCAATAA TGATTCTTGT AGCAAAAGCA ACCGTGAGCC ATTCGTGTA TGGAAATGAC	10020
ATAACAAGCA GTGACCTTGT TTATTTCTTT TCATGCTCGT TTATTTTGGG ATTAGGGCTA	10080
TATTTAGGAG GTTCCAGTGT ATGAGTTATC CAGAAGTTTA TATCATAGGA AGGCAAGTCG	10140
ATGGCGTTTA TGTTGAATAC TTACATGGAG CAGAGCAAGC CGATTTATTT TTCGATTATA	10200
CGATAGCTCG TGATGAAAGA AATCATATGA ATAAAACCAA TATGAAAGAT GGTACTTGGA	10260
AAATTTTGAA ATATGGTAGA CCGATAACGG TCGAATTTAA CAATAAAAAA GCCAACCAGC	10320
CAATGGTTGA CTAAGAAGAA TATTTTACCA GAAAAGTGGT AGCTTGTGAT ATGTGAGGTT	10380
ACTTTGCCCC AACATTGGT CACAATAAAA ATATTTTATC ATGAGTAAAG AAAGCTGCCA	10440
ACAAAAAAG CCGGATTTCT CCGACTGTGG ATAATGTGTG GATAACCTTG TGGAAAAGTC	10500
GTTATCATT TACCACAAAT CACAAACGGG GGAATCAAAG GATGGTACTT TTTGACGTAA	10560
AGAAATATGA AACACCGGAT GCAAAGGACG TAGATATGGA GCAAACTAAA CATAACGTCA	10620
GTGTGTTTCT GTCTGCCTAT CTTGCTGCTA GATGTCGTGT TGGCCAGCCG AGGGAACCAA	10680
AAGTAACAGC TTCATTCTCT TTGGTTCCAC CATCCACGGC CAAAAACACT TTTGAAGCCG	10740
AGCAAATGTT AATCCAGAAA GAAGAAGCCC AAGAAGAGTT TGATTATTTG CATAAGCTTT	10800
TTGTTAGAGG TTATTCTGCG ATTCAGCATC CGCACAAACC AGATGTTACC GAGCGAAGAA	10860
AAAGAATCTT CTATGATCGA TATATCAACG GTAATCCAAT CTATCTAGCA GCGCAACGAA	10920
ACTGCATCAG CGAAGAATCA GTGAAACAAG AATCTAACAT GATTATTGTT CAATTTGCTT	10980
CAGCGCTGGA ACTGGTTGCT TTTAAGTAGC CATTATTATC ACTTTTATA CCTCTTTTAT	11040
ACACTTTATC TACACTTCAT ATACCTTCTA AACGAGTTAT TATGATAGTG TCAAAAAAAT	11100
AAGAAATGCG ACACACTTAC ACAAATTTA AACGGAACGA TTGCCTACTT ATTTTTTTGA	11160
TTTGAGATTA CAAGGAAGTA AAAAAATTCT ACTTCTCTCG TTTAGTCACT TGTGATCTCA	11220
TTTAAATTCT CTCGCAACCC ACCAATTATA AAACATAAGA AGTGAGGTGA ATTCCTCTC	11280

TCTTTTCT	ACAGGTTTGC	GAGAGTTAAT	GGAGCATAGC	TTAATCGGCA	GAGCAGCGGT	11340
CTCCAAACC	GTTGGTATAG	GTTGAGTCC	TATTGTTCCA	GTAAGTGGA	TAAGCTGCTT	11400
AAATAAATA	GATCGTCAAT	AAATGTTCCG	ACAAACAAAT	TGGCGCTACT	ACCTTTCCT	11460
AGGGTGCAT	TTATATGCAG	TCCTTTTTGT	TTAAGTGTA	GTAATTTTT	CATTTTGAAA	11520
GGGGATACT	AAGACAATGC	GTGTATTAAT	TAGAAGTTCA	GCATCTGGTT	CAGAGTATTG	11580
GGATACCGA	GAAAAAGAA	ATGTGTTTGT	ACCTAAAGGT	CAAGAACCTG	ATTTTGAAGT	11640
TACTGAAAT	CCTGAATCAA	TGCTAAGTAA	AGAAGCTGAT	TTACATGTTG	GTGGATTACC	11700
AATCTACTGA	GGGAATGTAA	CGATTGATAC	TGATGGAATT	AAAGGCGAAC	GATTATTAAC	11760
AACTGCACT	GCTGCTGATG	ATGAAGAACA	AGATGAGCTT	GTTCCGCTCG	ATGATGAATC	11820
TCTGTATTA	GAAGAAATGA	ATGTAAAAGA	ATTGCGTGAA	TATGCAAAAC	GAAAAGGTAT	11880
TGAGATTCCA	AGTGCTGTAC	GTGCAAAGG	TGAAATTCTC	AATATTATTA	AAGAATCTGA	11940
ATAATGCCCT	ATTGCCAGTT	TGAAGGTTGC	TCTAATACAA	CAGAAAAAGG	AGCTTATTGT	12000
TCCGACATG	CTAGGAAGTC	AAGAAAAAG	AAAAAGCCAA	TCAATGTTTA	TCATCATGAT	12060
ACCAATCAT	TTTATCGAAC	AAAAGAATGG	CAAGATGTCG	CTGACTTTGT	CTATGAAAGA	12120
GAAGGTGTT	GCTGTCAAAG	ATGTGGCCGT	TTTGTATTTG	GAAGGCAAGC	GCATCGGCAT	12180
CATGTGATTC	CAATCAAGAA	GAACGAAATG	CTCAAACCTG	ATCCAAACAA	TATTCGTTTG	12240
TTGTGTCCGA	AGTGTCATGT	GATTGAAGAA	AATGAAGCAG	ATGAGAAAAA	AGTTTTTCCA	12300
TCTTATTTTA	AAAAATGAAG	CCCCCTATC	AAATCTGATT	CAAATTTTTT	GTCGGGGGAT	12360
AGGGTAGGGG	GCAGTCACGC	GTGTCGTTAG	GTCAAAAATT	TTAAAAATAA	AAGGGGGGTG	12420
TATAAATAT	GACCACAAAA	GCGCAACGCA	AAGCGATTAT	TGATGAAAAA	GTTAATCACG	12480
AAAAACGCG	AATTTTAGAA	ATTATGCGCA	AGTCTGATTT	ATACACTATT	ACTCTTGATC	12540
CATTGATTGA	ATCATACTTG	GATATTTTGT	AAGTTTACCA	ATACAAATAC	ATGCTGTGGA	12600
AAGAAAAAGG	ATTTCCTGAA	ACCCAAAAAA	CAACAAACAA	GGyTGGAGmT	AyTAACAATA	12660
GCAAGCATCC	ACTAGCGCAA	CAAGTCGAAG	TTTGGGCCGA	TArAAAAATG	AAAGCATTGG	12720
ATTATTAGG	ATTGACCAAT	AAGTCAAAAA	CAGGCAGACA	AATTACTGGT	GGTTCACAG	12780
CTAGAGCAGA	TGnaGAAATG	AAACGGCCAG	AAGAAAAGCC	TGTAAATGAA	TTGGCAGAAC	12840
ATCGGAAAAA	ATGGCGTAAA	AAGGCAGGGA	ATGAAACATG	ATTGAACCTG	GTGTAAATTA	12900
TGCTGATTTA	TTTGCGAAAG	AAGTTCGAAA	ACATCCTAAG	AAATATCCGA	AAACGGTTCCG	12960
TTAGCAATA	GATCGTTGGT	ATCGATGGAA	GAAACGAAAA	GATATTTGGT	TTGATGTTGA	13020
TCTGCAAAAT	GAAATGATGG	ACTGGGTAGA	ATCTTTTATT	GTTCaTACAA	AAGGCGATAT	13080
GSTAGGTAAA	CCATTTCTTT	TAGAGCCATG	GGAAAAATTC	ATTTATTCTT	GGATTTATGG	13140
CTGGGTAAAA	GAAAATGAAA	AAGGGCAAGT	AGTCCGTGTT	ACTCGTGAGG	CATACGTACA	13200
ATACCPAAG	AAAAATGGGA	AAACATTAAT	AGCCGTAGGG	GCGTTGGGAT	ATGCTATGTA	13260

TGGCGAAGGT GCCTTATCTG TCGATTGCTA TGCATGTGCT TCTGATTTTG CGCAAGCCCA	13320
GTATGCTGCT AAGCCTTTTG CCGCTACTAT CCTAAATAAT CCAGTGCTAT TAGATGGGAC	13380
TAAAATATTT AAAGGTCCAA AAGGCACCGT TTCAAGTATT ACGTATGACT ATTTACATGG	13440
AGATATGGCT TATACAAATA AGTTTATTGT TCAGACAAAA AACATTGATA ACATAGAAGG	13500
TTCCAATCCA TATTTTGTTT TAAATGATGA GCTGCATAAA CAAGAGAAAA TGGAACAGTA	13560
CGATAACTTT AAATCTGCAC AAATTTTATT GCCACAGCCG TTAATGTTTA ATATTCTAC	13620
AGCTGGTAAA GGAAGTAGTT CGGTTGGTAT TCGTGTTTAT AAAGAAGCAA AAGAAGTCTT	13680
GAAGCGCGAT GATAATGATT CAACTTTGT TTTGATTAT GAACCAAATA AGGGATACGA	13740
TTGGACaGAT AGAAAAGTTT GGGAAATGTG CAATCCTAAC TGGGGAATAT CTGTTGATT	13800
GTCTGCTTTA GAATCAGCCT TTAAGTGC GCAACGTTCC GCTCACTCGA AAGCTGAATT	13860
TTTaACGAAG CATTTAGATG TGTTTGTGAA TGGTGCAGAT AATTTCTTTG AACAAGATCa	13920
AGTGAACCG TGTTTGGTTC CCACAAATGA ATTAGGTAAC TTAAGTGGTG AGCCATGTTG	13980
GATTGGTTTG GACTTATCTA AAAGTCGAGA TTTAACTTGT GTATCATTAA ATTTTCCTAC	14040
ATGGGATGCC GAAGGAAAAG CGATACTCAA AGTAAAACAA TTATATTTTA TTCCTAGTGA	14100
AAATATTGAT TTTTCGAGAAA AGGAAGATAA TGTGCCATAT TCTGAATTAG CAGAACAAGG	14160
ATTTGTTGAA TTTTGCATG GTAAGTTAAT TGATCAAGAA CAAATATTTT ATTTTATTGA	14220
AGATTGCATG GATTTTTATG ATGTTCAACA AGTCAATTAT GATCCAGCGA TGAGTGACCG	14280
ATTAGTTGAA AAATTGGAAA ATTTAGGCTT GGAATGTGTG CAAGTTGATC AGTACGCAAG	14340
AGTATTGAAC TCGCCGCTTG AAGATGCCGA GCGATTATTT TATGAGCAAA GGATTATGTT	14400
TGATAATCCT TTATTTTTGT ATTGCGCTTT AAATGTGGTT GTCAAAATGG ATTTTCAAGG	14460
CCGTAAAGTA CCAAGTAAAA ACCAGTCAA GAGAAAGATT GATGGATTG TTGCTTTCCT	14520
TTGTGCGCAT AAGGAAACAA TGGATCAAAT GATTGATGTC AACGAAGATG ATATGGATGA	14580
ATATTTAGAT TCTATCTATC GATAATAGAG AGGCGGTGAG ATTTTGAAGC TAAGAGATAG	14640
ACTTTCAAAT GCAGTCTATG GATTTTGGGA AAAGCGTGGC TGGATTGAAG ATATTTATGG	14700
CAATGTAACA AGATATTCAC AACGTTTTGT TAACGATTCT TCTATTATGG AATCGTCTGA	14760
TGTTTATGAA TTGGTACAAG ATATTCTAA TCAAGTTGCA CTAGCAGAAC CAGTAGTAAT	14820
TGGGCCTGAT GCGAGGAAG TGAAAAATCA TTTCTTGTTA AACATATTGA AAAATCCTAA	14880
TGATTATTTA ACTGGTTTTG AATTGCAAAA GCTTGAAACA AATACATTGT TAATCAATGG	14940
AGAAGCTTTT CCTATTACAG ATAATGACCA GTTACATTTG GGATATGGTG TTCAAACGAA	15000
ATTAGATGAT CGTTTGATTG AAAGATTTTC AATGAATGGC CAACCAATAC CAGGGAGTAT	15060
GATTCGTCAT ATAAAAATA TTGGTGTGGA TTCCTTAAAA GGTGCTGGAA TTATTGATCT	15120
TGCAAAAAGC ACACTAGAAG GTGTTTTAAG TGCTGAAAAG GTTTTGACAG AAAAATATAA	15180
GAAAGCGGC TTGCTCGCTT TCTTGTTAAA GCTGGATGCG CATATCAATC CAAATAACAG	15240

CGCTCAACAA AAGTAGTAA AAGCTATTT AAATCAGTTG GAAGAAACGC AAGATAATGA	15300
TAGTCATTCA GTTATATGA TTCCTTTGGG CAAAGSATAC TCAATTGATA CTTTAAAAG	15360
CCCAATTGAT SATSCAGCTA TTCCTTAATTA TTtGGGTGTT CACAAPAAAG ACCTAGGAAA	15420
tTTTTAGGAA TAGTGTAAA TACTTATCAG CATTATAGAG AACaGaTATT GnAAAAGCAa	15480
TGATGTATCT GCACACAAA GCAATTAAmC CAATATTAA AATAAGAGC TAGCATTACT	15540
CGGCTCTTTT TTTTATGCTT AATTCGTGT ATCGASTGGA ATGGAAAATT AATATTTTGG	15600
ACTTTGTACC TTATTCACC AAACAAATA TTGGGTACAA CATTGTTCGA ACTGGTATTA	15660
CCAGTCCTGA TAAATGGCA GAAATGCTTG GTTTCTCTAG ACAAATACT GAAGCAACAC	15720
AAGCCGTCTA TATTTCAAAI GATTTAACGG AAATCGGCAA AAAGAATGCT ACCGATAACT	15780
CATTGACAAC AGAGATGAC TTGAAGGSAG GTGGTAAGAA TGAAGAAACA GGAAATTCGG	15840
ACATTTGACA TCACAAACCT TAAACAAGA AGCGAAGAAG ATAGTCAAAC ACAGATTGTT	15900
ACTGGTTATG CGGCGGTGT TATAGTCCA ACAGAATTAT GGAAGGCCCT AATGAAGTG	15960
ATTAAGCCTG SAGCTTCAG TCGTCTTTG TCAAATCTG ATGTTCTTG TTTATTCGAT	16020
CATGACTGGG GCATAGTAT ASGGCGCACA AGAAGTGGA CTTTGAACT TGAAGAAGAT	16080
GATAAGGGAC TACGATTGA AGTTSAGTTC CCCCATACA ACTGTTGCCA ATGACTKGAT	16140
TCAATCAATG TCACCTGGGG CATTATCA GTGTAGCTTT GGTTTTTATC CaACGGAAGA	16200
AACTTGGGAT TATGTTTAC AGCCGTTTT AAGAACTATC CaTGAAGTCS AATTGTATGA	16260
AGTTTCTATT GTTCTTAC CTGCTTATGA AGATACGAA SCAGCACTAT CAAGAAACAA	16320
ACAAGAAATG AAGCAAGATA TTAATCTAG AAAAATTA ATTGAAAAA TCAAAACAGC	16380
GCTTGAAGCG TAGGAGGAAT TTATTATGAA CAAAGATTA TTGCGTCAAT JACAAGCTCG	16440
TCACGAGAAA CGATTAGTG ATTTACAGG CAAAATTGAA TCTGGAGAAG TCGTGAAGC	16500
AGATTTAGAT TCAGTTAATG AAGAAATTGA TGGTTTATC GATGAATTAA AAGCCATTAA	16560
AGCTGAATTA GGGGTGATA ATTCAGATC TGGTGTGGT GAAGGCGATG ACGGGACCGC	16620
TAAATCTGAT AATACTGAT ATGAAGTAA ACAAGATCGT GAGAAAGATA CGAACGGAAA	16680
CAACGATGAT AAACAGAAAG AAAATCGTGG CGGCATGATT AGCCAAGAAC AGCGTGATGG	16740
TTTGTTACGC ACATTCATG AAGCATGGA GGCTAGAAAT GTGATGTCTA ATGAACAACG	16800
TGAAAAACAA ATTCGTAAAG CATTGCTGA TTTTGTTATT GGTAATATTT CAGAAAGTGA	16860
AGCACGTGCA TTAGGTATG AAACAGGCAA TGGTTCAGTG ACAGTACCAG AAGTGATTGC	16920
ATCCGAAGTG ATTCTTATG CTCAAGAGA AACCTATTG CGTAAATACG GAAGTGTGAT	16980
TCGCACGGCT GGCAATGTA AGTATCCAAT TCTTGIGAAA AAGCAGAGG CTAATGTAAA	17040
CAAAAAAGAA CGTACGACAG ATATTCTGA AACAGCGATT CAATTTGATG AAATTCTACT	17100
TGATCCAGCA GAATTTGATG CATTCGCAAC TGTAACGAAA AAATATTAA AATGTCTGG	17160
TGTGCCAGTA GAAGTATTG TTGTAGAGA ATTGAAPAA GCATATGTT CCAAAGAAAT	17220

TAATTATATG TTTAATGGCG ACGATGCAGG aAATGAAAAC CCAGGAGCTT TAGCTAAAAA	17280
AGCTGTTGCA TTTGAAAAAC CTGTAGATTT AACAGCTGCA GGTGCTGGTC AAAAATTATA	17340
TGATGCATTG ATTGAATTTA AAAATACACC AGTAACAGAA GTAATGAAAA AAGGTCGTTT	17400
TATTATTAAT CGTGCAGCTT TAACTGCTAT TGAAAAAATG AAAACAGATG ATGGATTCCC	17460
ATTGTTACGA CCATTCACGC AAGCAGAAGG TGGTATTGGT TATCAATTAG TTGGTTATCC	17520
AGTTGATTGG ACCGATGCAG CAGATAAAAA AGGTGAACCA GATACACCAG TATTATATTT	17580
TGGTGATTTT TCTGCTTTCA AAATTCAAGA AGTTATTGGA GCGTTAGAAA TTCAAAAATT	17640
GGTTGAAAAA TTCTCTGGAA AAAATCAAGT TGGTTTCCAA ATTTACAAC TGTTAGATGG	17700
CCAATTAGTT TACTCACCAT TCGAGCCGGC TGTTTATCGT TATGAAATTA CAAAACCACT	17760
AGGTGGTTAA TGTGAATAAC GAaGCTGAAA CATTATCTTT AGAAGAAAAA TTCAAAGCAC	17820
ATATTCATTT TGAAGAGGGG ATGGATGATT CCAaGCTCTC TTTTtATTTA AATATGGCAA	17880
AAAATTATGT AAAAaCTGCa ACTGGAGGGC AAGAAGAATA TTTtAATTTTG ATGGTTGCTG	17940
GTATgCTTAT GAaTATCGTG TTTCAGAAGA TGAATTAGAT AAGGCGTTGA ATGCGATCAC	18000
GCCATTTATC ATCCAAGGAG TGATTCAACA TGCCGAAGAG GCAGACGAAT AGGTTTCGAT	18060
GGAAAGCGGA CTTGCTAAAT GTAAAAGAAG AAACAGATTC GAACGATAAA GTAGTTACGA	18120
CCTATAAACT TAATAGGCTT TTATGGTACG AAGATATTGG AGTAACTGCA CAAGAAAAAT	18180
ATCTTTCACA GCAAGCCAAA ACAGACGTTG TCAGACGGAT TAAAGTGAGA TTGGATAAAT	18240
CTATCACAGA AAAGTTTAGC GCTGTTAGAA TCGATTCTGT GACCTATAAA ATCACTCGTA	18300
TTTACACAAA TATGGATAAA CGAGAAATGG AGTTGAGTTT GGCTTATGTC GATTAGTTTT	18360
GAAAAATTGA GAATAACTCT GAAATCAGTA GGCCTACCTG TGACACGTGA TAAAGCGGAA	18420
AAAGGAACGG ACTATCCATA TATTGTTTAT TCAAATGTTA GTCAAGGTAA AAAAATGGCA	18480
TCGTCTAAAG TGCATAGACG AATGCCCTAC TATCAAATCT CTTTTTATAC AACAGGTACT	18540
GAAAAGGATT TAATTGCTTT AGAAAATGCA TTGGAGGAAG CTGGTATTCC TTACACTGAT	18600
TTTGTAGGCA TTCAAGGCGA TGAAAATGAT GATACTGTGA CAAATTTTTA CACATATGTG	18660
AGGTGTATTG AAGATGGAAA ATAATAATGG TTTTGCAGAT ATGGCAGACT ATTTAGGAAA	18720
GCTTTCGCAA GTAGATGCGA CAAAATTATC AATAGAATCA TTAACCGCTG CAGCTAATTT	18780
TTATATGGAA AAATTACTAC CCAATATACC TAAATCGCTT CTAAAAAGA AGCACATGGT	18840
TGATCAAGTG AAAGTAAATA TTAAAGATAA TGAAGTACAA GTAGCTTTTG AAGATACAGC	18900
TTTTTATTGG CGATTGCTG AAAATGGAAC AGTAAATCAA AAAGCGCAAC ATTTTGCCAG	18960
TGGAACATTT GAACAAAATA AAGATCAAAT TGAAAAAATT ATGACTCAAC AAATATTAGA	19020
TTTATGGAaA GGATGAGTAA ATTGGGAAAA CAAGATGTGT ATTATTTTGA AGGCTTAGAT	19080
GACATCTTAA TTGCCATGAT GACCACAAAA GATGCAGTAG GTACAGAACC AGCATTTGGC	19140
GAGGTTGTTC GTTTGCCAAT AGCCACAAAA TTGGGAATTA AAGGAAATGG AACAGCTTTA	19200

GAAAAATGGG CATCAAGTAA AATGTTCCGA CGCGTAAGTC GTGAAACGAA ACATGAAATT	19260
GCGCTAGATC ATGTGGGCAT TCCTATTGCG GTGATGGATG AAATAAAAGG ATTAATCGCT	19320
CAAAGTGGAG TGA CTTTGG TAAAAACACT GCGCGAGAAT TTCCTTATTT TGCCTTTGGG	19380
TTTATCGGAA ATATTGAAAA TGGTGGAAAA AAAGCTGTTT GGTATCCTAA AACGCAGTTA	19440
TCAAATGTTA TTGATGAAGA ATACACTACT GCAGAAGATG AAACCAAAAT TGATGATGTA	19500
ACTGCTAACT TTGTTTCAAC TGGTTTAAACA TACAATAATG TTATGTATTC AAGTTTTGAT	19560
TCTAATCGGG ATAGTGCTTC AATAGAACTA TTTGAAAAAT TTATCGCACA ACCTGTTTAT	19620
GACGAAGAAC AATGGAAGAA ACTAGCAGGT CCTTCACGTG GAGGTGGCAG TGAATAATGG	19680
CAAAATTAGC GGATTATGGG ATTGTTATTT CAGATACACC CACAGTTACT ATTCAAGGGC	19740
ATCAGTTTCC CATTTTGTTA ACGATGGAAA CAATGGAGTA TATTGCTGAT GTTTATGATG	19800
ATGACTATTC AAAATTGAA GCAGATATGA ATGAAATGAT TAACAAGAGT GGTGGTCGTA	19860
TTTCATCAAA AGATTTATCA GCTTCTGATT TAAAAATTAT GCGTGCATTG ATTTACGGCA	19920
TGCTAAGAAC TGGTGGATTG GAAGAAACGC CAGAAACCAT TTTTAAGTTT TTAGGTATGA	19980
GTGCCACAAT TGTTGAAATT TATGGTGCAT GTATGGAAAT ATTTGCAAAG CAGAATTTTC	20040
AAGTTGAAGA CTTAAAAAAA TCCAAGAAGC CACAAGATTA TCAAACCTCCG AAAAAAAGGA	20100
AAAACAAAA GAAAAAGCCT CAACGGAAAT AGGAACGCCG TGGGCTTTTT ATTTATATGT	20160
AGCTCTCACT CTTTTGGGAT GGAGTGAGGA CTTTTTTTTA AAAGCAACTC CCAACTTG TG	20220
GCTTAAGTCA TATATTCAGT GGTAGTAAG TAATACGGAG TTTGAACCAC CCAGAAGTGT	20280
GACAATGGAT AAAAGTCCTT GGTGGTAGGA AAGGAGCGCT AACGTGTCAA AACAGAATC	20340
CGATGTTGTC TTAAATTTTA AGATGAATGG AGAAATAAAC TATTCACGAA CAATTAAAGA	20400
CATTAATAAA GAAATGAACT TAGCGGCTAC CGAGTACAAA AACCAGGTAT CCGCAATGGA	20460
TAAAAATGCA ACTCAAAC TG AAAAAATTAAC AGCTACAAAG AAAAAATTAG AAAAACAATT	20520
ATCTTTAGCT GAACAAAGAA CAAAATTATT ACGTGAGGAA TACGAAAAAT CAGTAAAGA	20580
AACTGGTGAA TATTCAGAGC AATCACA AAA GCTGTATAAG CGTTTATTGG AATCCGAAAC	20640
AGGTGAAAT AACTGCGTT CTGCATTGCA AAGTACCAAT GAAGCTTTGA AAGAGCAAGG	20700
TAATTTATCA ATAAAAACAG CTGAAAAACT AGCCAAAATT GAAAAAGCTG GAGACAAAAT	20760
TAAATCAGTT GGGCAAAAAC TGTCTGTTGG ATTAACAGCA CCAATTATGG GAATTGGTGC	20820
TGCTTCTATT GCCGCATTCA AAGAATTAGA TGAATGTTTG GATAATATAA CAACAGCAAC	20880
AGGGGCTACT GGTAGTCAGC TAGAATCTTT ACAAGCCAGT TTTAAACAG TAGCAGGTCA	20940
AATACCCGCG GATATGCAAG ATATATCAAC TGGTATTGGT GAAGTAAATA CTCAATTTGG	21000
CTTAATGGAT AAGCAATTGG AAGATACAAC AGGCCGAATG CTTAAATTTT CAGAAATTAA	21060
TGGCTCAGAT GTTCTCAAT CAACTATTAA TGCAAAAAA TCAATGGACC TTTT TAGGTT	21120
GTCTATTGAG GATTTGCCAA TGATTTTAGa TTCAGTATCT AAAACTAGCC AAGATACTGG	21180

AGTAGGGGTA GATCaGTTAT TTGATGCCGT AAAtAGAGGT GCGCCCCAAC TCAAAGCTAT	21240
GGGACTTGGT TTTTCTGAAT CAACTACGTT AATAGGTCAA ATGGAAAAAG CTGGTATTGA	21300
TTCAGCAGGA ACTCTTGGCT ATTTGGCAAA AGCTAGTGTC GTATATGCGA AAGATAATAA	21360
AACCATGCAA GAAGGGCTTA GCGGAACTAT TGAATCTATT AAAGGGGCCA CAACTGAACA	21420
AGAAAACTC ACTATTGCTA GTGAAGTTT TGGAACTAAA GCTGCTTCAA AAATGGTAGA	21480
AGCTATTGAT AGCGGAqCGT TGTCAATGGA TGGTTTAGCA GATTCAGCAA AAAACGCAGC	21540
TGGCACTGTG GATCAGACGT TTAGTGATAT TCTTGATCCA ATTGACCAAG CAAAGTTGGC	21600
ACAAAATCAA TTTAAAATAG CAATGGGTGA ACTTGGAGAA CAAGTACAAA TAGCATTATT	21660
ACCAGCATTT CAAGCTGCAA CGGATGCAAT AAAAAAAGTT TCAGAATGGT TTGGAAGTTT	21720
AACAGATAGT CAAAAGCAAA CCATACTGAA AATAGCTGGT GTTGTGGCTG CTATCGGTCC	21780
AGTATTAGTA GTTTTAGGAA CACTTGCTAG TTCCATTAGT AGTTTGATTC CAGTTATTGC	21840
TTTTATTGCG TCGCCAATTG GTTTAGTAAT TGCGGCGGTT GCCGCTTGGG tAGCTGCAAT	21900
CGTAGTTGCA TATAATAAAA TCGGTTGGTT TAGGGATTTT ATCAATACCT CCTTTAAAGT	21960
AATTAAAGAT ATTGTGGTTG GTGTATTTAA TGTTTTGAAA GATACGACAA AATCTACTTT	22020
TGATTTCATC ACAGGATTTA TTGGTGGTGC CATGGATGGG GCTGCAAAAA TTATTGGCGA	22080
TTATGTAAAT GAAATTAAGC GTATTTTTGG CGGTATCGTT GATTTTGTAa CGGGAGTATT	22140
TACTGGAGAC TGGTCAAGAG CGTGGCAAGG TGTTGTTGAC ATTTTGGTG GTATTTTGA	22200
AGGCATCGCT GCAGTAGCTA AAGCTCCAAT CAATGCCATG ATTACGTAA TCAATGGATT	22260
TATTGGTGGA TTAAACAATA TAAAAATACC TAAATGGGTG CCAGGAATTG GCGGTAAAGG	22320
ATTTCATATT GGAAAAATCC CTTATTTAGC AGAAGGTGGA ACTATTCTAA ATGGCCAAGC	22380
CATTGTTGGT GAAGCTGGTC CTGAACATT AACCGCTAAA AACGGCAAGA CAACAGTAAC	22440
TCCATTGTCA CCAGAAGAAA AAGCTCGTGG AATTGGTGGT GCTTTGAAAG GTGGCAACAC	22500
TATTGAGCAA CATGTTTATA TTGGCCAAGT AGATGCAAAT AATCCGAGTG AGTTAGATCG	22560
AATGAATCGC AAGCTTTATA AAGCAAGTGC GCAAGCTTTC TATGACTTaG GAGGTGTTCC	22620
aACGTGATTT TTATGAATCC TGATGAACCA AATTTCATTT GGAAAGATTT GAATGCAGTT	22680
CGTGATATGG GGTGCATTAT CGAAAATGAG CTGTCAGAGG TTTTACCAA TAAACGATAT	22740
GAAACGTATT CGATTATCGG AAGAAGTGGT GAATTTAATG AAACGTTCAA TGATTATGAA	22800
CCCTTTGATT ATGAAATTGA AGATGTAAGT ATTCCATATG AAAATTTAGC GGCAGTCAAA	22860
AGATGGTTAA CTGGTAAAAG TAACTTATT ACTCACAATG ATGAAGATAA ATATTTAGAT	22920
GCTATTTGTA CAATGAGTAA ACCAACTTCA TTCAAAAATG AATGGGGTGT TTTTATACC	22980
TTTAACGTTG AATTAGATG TCAACCGTTC AAAAGAAAAG TAAACGAACA ACCAAAAGTG	23040
ATTAAAACAA AATCAATTGA AATTACTGAT CACGGTGATG AAATTGCTTT TCCTTATATC	23100
GAAATTAATT CAAAAGGTGG CGATATTACG TTAAACATTG GTAGTAACTC ACTAACGATT	23160



TTGCGTACAC AATCAGGAAT CGTCACTATT GATACCGAAA AGGGAAGC AATACAAGAA	23220
GGAAATCCAC TATTACACG CGGCAGTTGG ATAAAAACGA ATCCCGGTCA AAATAAATTA	23280
AATATATCAG GAAATTTTAT AGAAGCTAAG TTTTGAATA GGAGCGCGTA TTTATGACAC	23340
AAAATTTTAT TTATGCCTAT ACTGCTATTC CTGAAAATTT AAACGATAAC GGAATGGCTT	23400
TGCCAGATTG GCAAGATTTA CCAGAAATTA ACCGTGTGTT AAATGGTGTG TATCGATTCT	23460
ATGGTAACTA TGCAAGAGAT GGCCAGTATC GCTCATACTT AAAAAAGGA AACTTTCTAA	23520
AGGCACAAGT TGAAGATGGA TCATATCAAT ATTTTGAGAT TTACAATATT AAAAAAATC	23580
TGCAGTCAGT TTCAGTGACA GCGAGACACA TTGGTTTTAT GGCAAATAAG AATTTTATTA	23640
TTGATTCGTT CACTGCTAAC GGAAATGGCA CGCAAATTAT GAATAATTTA AAGGCTGCAT	23700
TAACGTTTAA GCAACGGTTT AACTATTTGT CGAATGTCGG TACTACACAT CAATTTACAG	23760
CAAAACAAGT AGGTCCAATC GATGCAATTA TTGGTTCTAA CAATGGCAAT CAAAATTTAA	23820
CAGGTGTTAC TGGTGGAGAA TTAGAGATGG ATAACTTTAA TTTGAAATTA GTAAAACAAA	23880
TTGGAGCAGA TAATGGCTTT AGAATTGATT TTGGAATTAA TTTGGAAGCT ATAGATGAGG	23940
ACTATGACGA CGAATCAATT ATAAACAGTC TTTTCTTAT CGGTGGCGTA CCAGACAATG	24000
ATTATGACCA AGATAAAGAG CCAATCACGT ATGGCTTTTT AGAAATTGCT GGTGTAAATG	24060
ATAGTAACAG AAGAATTGGA AAACGTGAAA ATTCGGAATG TAAACAGTT GATGAGCTTA	24120
AAAAATGGGG CCAGTCATTG TTTGATAAAG ACCGCATTCA TGAACCAAAA GTAACGCACA	24180
CTATTAGCAT GGTAGCATTG GAACACACTT TGGAGTATGA AGACATGTAC GAAGAGCTTT	24240
CTTCTTTGCA TTTTGGTGAT GTAGTACATG TTAGAGCAAA AGAAGTCGAT ATTGAAGTAA	24300
CGGAGCGCAT GGTGGAATAT ACTTGGTTTC CGACTTTAGG CAAATTTAAA AATATTGTTT	24360
TGGGGAATGA TTTATCACTT TACACCTCAA CAGTAAATAA TCAAACCTCAA GAGCTAAAC	24420
AAAAAATTGA TAATCGAACA GAAACATTAG TACAAAATGT TTTAAATGCA ACGGCATGGA	24480
TTACTGGAAA CAGTGGTGGG CATGTCGTTT TTCGTCCAGA AAAGGCCCGG TCTGAAATTC	24540
TTATCATGGA CAAAAACAAA GTAGCTACCG CAAAACGTGT GTGGCGCTGG AACTTAAATG	24600
GTTTGGGTTA TTCCTCCAAC GGCGTGAATG GTCCGTTCGA ACTTGCTATG ACTTCTAAGG	24660
GAGAAATCGT TGCTGATTTT ATTAAAGTGG GCATTATTAA CGCGAATGTT TTACAAACAA	24720
GCTTTAATAA AGCAACAGAC GATGTACTAA AATTAGTAGC TGGTGCTTTG CAAATTTGGA	24780
ACAATAAGAA AAAAATCATG GAATTGACTA AAAAAGGGAT GGAATTTTGG AATGGCTCTA	24840
GTCATATTGG GACAATTGGT ACGAAAGGAA ATCCCTTTCC AGGGGTAGTA GATAAAAATG	24900
GAAATCCTGT AGTTTCTGAT GGGAAATTCAT TGCTATTAGT CGCAGATAAT CCCAAAAAA	24960
TTATTGGTTT GTCTAACCAA TCAGGAACAG GACATTTAAT TACTGGTCCT ACACAGTTTT	25020
TTGTTGGAAA TAATTTTAAC TTTTTGGTC CGAATGGAAG TAAAGCAATT CTGACAGTTG	25080
ATCGATTGAT TGTGGGCGGC AAAGAAGTTA TACCTGGTCA AAATGGTGGT GGCGGTTCTG	25140

GAGCTGGAAC AGGTGGTTAT CCATCAGAAG TTACAAGCGA TGCAGATAAA TTTGCTTGGG	25200
ACTTATGGAG TTACCTATTA GCTAACGGAT ACAGCAAAGC AGCTGCTGCA GGTATCCTyG	25260
GAAATGTACA AGGAGAAGTT GGTCCAAGTA TGAACCCAGA TACCGAACAA ATAGGCGGTC	25320
CAGCTTACGG ATGGGTTCAA TGGGACGGTT CAGCATATCC ATTGGTAGGC GCACCAACTT	25380
GGAATGGCCG AGAATATGTA CAACGCTTAA TCGCAGCTGC AGGTATCAAA CAAGACTATA	25440
GGACGTCATT AGCCCAAGCT CAATTAATTA ATTGGTGTAT GTTCAATGGG CAATGGTTAG	25500
GACAAGTAAG TCCATTAACa GTTGATGaAT TTaAAnnTCT ACGGTCTTTG GTTGGTAACG	25560
TTACTTCTGT TTTTTCAGAA	25580

## (2) INFORMATION FOR SEQ ID NO: 20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1260 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ACTTAAATAC AAATGTTGGT CGTGCACAAT GGGAACCTCA AAGTATCGAA GCGCCCAAAA	60
ATACGTTTAT TACTGGTAAA TTAGCCAGTG CAGGTGTCTC AGGTTTTGAA CCTTTGGATG	120
AACAAACAGC TACTCGTTGG CACAAAACAA ACATTACAAC AGGTCCCCTT GACATCACTT	180
GGAACCTAAC TGCCCAACAT AGAACTGCTT CTTGGGATTA CTATATTACT AAAAATGGCT	240
GGAATCCCAA TCAACCATTA GACATTAAAA ACTTCGACAA AATTGCTTCA ATCGACGGCA	300
AACAAGAAGT TCCTAATAAA GTTGTTAAAC AAACAATTAA TATTCCGACA GACCGCAAAG	360
GTTATCATGT CATTTACGCT GTCTGGGGCA TTGGTGATAC GGTGAACGCC TTTTACCAAG	420
CGATTGATGT GAACATTGAG TAAAAATACT CACATCTTAA ACTAAATTAA AGGAGTTCTC	480
AAAAAATGAA GCTAAAAAAA ATAATTCCTG CTTTTCCCTT TCTTTCAACC GTTGCAGTTG	540
GCTTGTGGTT AACGCCTACT CAAGCTTCTG CAGATGCTGC GGATACGATG GTAGATATCT	600
CTGGCAAAAA AGTGTGTTGGT GGATATTGGC ATAAGTGGG CTCAAAGGA CGCGATGGTT	660
ACAAACAAGG AACATCAGCA TCACTAAACC TTTCAGAAGT AAATCAAGCC TACAATGTGC	720
TACCGGTTTC CTTGATGAAA AGCGATGGCA CGACACGGAT TCCTACGTTC AAGCCTTATA	780
ACCAAACGGA CACTGCCTTC CGACAAGAAG TCGCACAATT AAATAGTCAA GGTGCGCGAG	840
TTTTATTGGC ACTTGGTGGG GCAGATGCAC ATATTCAATT AGTCAAAGGC GATGAACAAG	900
CCTTTGCGAA TGAAATCATT CGTCAAGTGG AAACATACGG CTTTGATGGT TTAGACATCG	960
ACTTAGAGCA ATTGGCGATT ACTGCTGGCG ACAACCAAAC CGTCATCCCT GCTACGTTGA	1020
AAATAGTCAA AGACCATTAT CGAGCACAAG GAAAAAATTT CATCATTACG ATGGCACCAG	1080
AATTCCCTTA TTTAAACCT GGTGCCGCTT ATGAAACATA CTTACTTCC CTAAATGGTT	1140
ATTATGATTA CATTGCCCCA CAATTATATA ACCAAGGCGG CGACGGtGTC TGGGttGATG	1200

AAGTTATGAC TTGGGTTGCT CAAAGCAACG ATGCTCTAAA ATACGAGTTC CTCTATnATA 1260

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8188 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TCCAATGGCT TATCGATTAC AACAACTGGA AGTtCTTTTA AATCAAGCTT GCTTTTTTCCA	60
ATTAGAtACC TATGGAAACC ATCAwTAATA ATATATTCGT CTTTTTTATC ATCACAATAG	120
CAAACAATCG GTTGTGTATA TCCGTCTTTT AAAATCGATA ATTCTAAAAG CTCCATTTCA	180
GGTCCAGCGA CATGATTTGG GTTATAACTA TTACTTTTTA TTTTCTCTAC TGGAACCATT	240
TTTACATTTA AAACGGGCAT ATCAATCTTG TTCATTGATT TCTCTCCTAA AAAATAAATA	300
ATTCTTTGTC TCTCTATATT GTTTAAACA ATAATGTTTC AGATTATTTA AGTACTGGTC	360
TTTTTTTAAA GAAATTCGTA CACACTTTTT TTCACATTTC TTGTCTATGA AAAATAAAAC	420
CTTCTCCAAT AAATACTTTC GAGTATCGTC ATTTTCATCT TCAACAAAAA CATCTCCAAT	480
ATCAATAAAC TCATTGGAAA TTTGAATGGA AGAAAATCCA ATTATCCTTT CCTTATTCAT	540
TAACAAAAAC CAAATAGTAC TAGATGTATT TATTAAATAA GGTAATTTTC TTCTATATAC	600
TCTTTCAGCA AAATATTTCC CCATTAAACT GTAAAAATTT TGTyTATTGT ATGGGGGTAT	660
ATAAACTTTA ATATCTATTT TGCTCATTTTC ATTCCTCCTA TTTTTTGAC AATATTTTCG	720
TATTTCTTTC CTTTAAATTG ATTTATCCCC AAAAAGTCTT TATATTCTTT AAACATTTCA	780
TATAGTTCAG CTTCATCTTT TTTCGTAAAC CTATAATTTT TAGTAGATAA AACAAAGTCA	840
TTTTTTTCTA AAGCTTTAGC TATCCCTTTC CAGTCATTCC AAAGCCTTtC ATCTTTTATC	900
CAATCTTTTC TTTTtGATTC ATCTGGCATA TCTTGGATAT CCATATtATA tGTATtCktG	960
TAATAaGCAA AAaGAATACA TATCTTTCTG TAATAATGAT TCATTAAAAAG AGGAGAATAT	1020
AATCCTAATG TTTCTAGTAA AAAAATAGAA TATCTTTGCC AAGTCATAAA TGAAGGTTTT	1080
TGAGAAGTAT TTAATCCTAA AAGACTAGTT TTACCATATA TATTTCCAAA ATTAActCCA	1140
CTTACTCTTT TTAATAATTT ATCCCATGTT TCaGTTCAA GTTCAGCAAA TTGATTTAAC	1200
GCTTTTCTTT GTTGAATCC GAAAGGCTGG CATACTCTCA TGTCCCGGTA CTTATATCCA	1260
ATCTTATACA TTTTCTCATA GATATAATTA TATTCTAGTT CACATTTAAA AGTTGCAGCC	1320
CAAATGTCAC CAACTTTCCA ATCAAATAAA GGATAAAACA AAATATTTCC TTCTAAAGTT	1380
ATATTTATCC AACATCGATT CTTATATTTA TTTTCACCTT TTGTAACAGC ACTATGTCTG	1440
TGTAAACTTT CTTGTGCCCT TATACCTACA CCACATGATA TAGCACTTTG GGGATATTGT	1500
TCTTTATACC AATTTTGAAA ATAGCGAAGA AATCTGTCGG GATTACCATT TGATAATGAA	1560

TAATATTTTT	TTAAACCGTC	ATCAATATTT	TCCAAAGTTA	TGGCATTTTT	AGGAAGAGAA	1620
TGAATCCATT	TTTCTTTTTC	GCTTGGATCC	CACATTTTCC	ATTGAGGTTG	AAAAAACTA	1680
GTGTTGTTAT	CCTCAAAAAA	AGGTAAGCAA	AAATGATAAA	TATTTTTTAC	TCTAGGTAAC	1740
TGTTCAATTT	TTTTTATAAA	ATCAACAGTT	GCTGTGTAAT	TTGCTTCAAT	ATTTAATATT	1800
AATAGATCAA	AATTTCTATC	TAATTTTTCT	GCAACTAGAT	TAGCTAACTG	AACCATTAGC	1860
CCACTATCCT	TACCGCCACT	AAATGAAAAA	TACACCTTAT	CAAATTCTTC	AAATATGTAT	1920
TCTAATCTCT	TTAAAGTAGC	ATCAAAGACA	TTCATTTTCC	ATCCTCCCTA	AAAAATAAGC	1980
ATTTAATATC	AATGGTAATA	CATTAAAAAC	TTTGTCAATA	ATAAGGATAT	ATTCCAAAAT	2040
CCATTCAATTG	GAAATAGAGA	CCTTCTTTAT	TAATGTATTA	TTACTTACAT	AGCGAAgcAA	2100
TaCAAACATT	ATCCAATAGT	TTCATCTACA	AAAAATAAAT	TACTTTTAAA	ACATCTTGAT	2160
TATGAATTCA	TTTCCCAAAG	GCCTTTTAAA	ATTGATACAG	ATGTTCGTTT	ACTTTTTTTA	2220
GACTTATTGT	ACAAGTCTTT	TTAATAGTCC	CAGAACAAAA	TAGTTTGTTT	ATCCAGCTTC	2280
AAATCATTAA	AGGCTAATGA	ACGGTTTATT	TTTTCGATTA	CTTTAGATAA	AGCTACAGAA	2340
ATTTTATAT	TTTCACAATG	CTCAGTTTGA	ATTTCAATAA	TTATATCATC	AATTAGTTTT	2400
TGAATTTCTT	ACCTATCAAC	ATATTCAGTC	TTAGAAAGAG	AATCTTTTAA	AGCTACTAGT	2460
GATTCAACGA	TTTCAGACAT	TTTCTTCCTC	CTTTCTAGTG	GCCTGTAAAA	CCGTGTAGAA	2520
CAGAGTCCAT	CGAACCGGTA	ATTACACTAT	CTACaGCGCC	ATTCCAATTT	ACTCTAcATT	2580
TTCAATTTTC	AGATTTGTCA	GCCCATACAG	ATAGCCTTTT	TGACGTTTCAT	ATTCGGATTA	2640
TACCATTGAG	CCGCAGAAGC	AGATTCGCTC	GACACAAACA	ATCCCATACC	AACAAAACCA	2700
ATACACATCA	CACTATAGCC	AACTAGTTTT	TTCTCTCGk	ATCCATCTCC	TAAAAAATAG	2760
AATAGTtCTT	GTAATTATTC	ATAAGATATA	CTTTTCCGTT	TTTATCCTT	TCTGCTTTAT	2820
AACAACTTTT	TTCAAGTAAA	GAATGACATA	CTTCTCTTA	AAAAAACAGT	CTGTTTTTTT	2880
CTATTTTTTC	GGTGTATAAT	TAATACGAAT	GTTCAATTTT	TGTTTTAATA	CGCACTCCGT	2940
ACATAATAGA	TCTTGATTGC	ATCAATCTTT	TGTACTTTTT	GATTTATCTT	AAGCTGAATG	3000
AATAGAAACC	TAGTTACCGC	ATTACTTTTT	TGTTCTCCAA	ATAACAAGTA	ATGATCCGTT	3060
TGGGTTATAT	CTTGTGAAAC	ATTATGAATA	CGACCAATGT	TTTTTTGTTG	TACATGTGGA	3120
TCGTTGTCAA	TTCCATTTTC	TTTTATACAA	CAATCGATTA	AATAGTTTTT	TACGGATTGA	3180
TTTCGTTTCGT	ATATGTCCGT	GTAATTTACC	CATTTTCTTC	CAAAATCCTT	TAACAAATGT	3240
TGAATTACTT	CAAATATGTT	GGATTTTATT	TTGTTTTCCT	GATTTCTTTT	ACTTTCTATA	3300
CTTAAACTTT	CTGCTTGCAT	CATCGTTTCT	CTTGATATTG	AGGAACTGC	TGTTTTTTTC	3360
TTCAGAAGAA	ACCCACCTAC	AACTTCTAAG	GACAAACAAA	AGCAACCAAG	AAAAACAAT	3420
ACTTTTTTCAC	AATTGTTTCA	TCTCCCTTTT	CTACTAAAAA	TTTTCATTGA	GAGAATTCTT	3480
TTATTTGGCT	TTACGAATCA	AACTTCGAAT	TTTCGTTTGG	TTATAGGTGC	GTTGATACAG	3540

TGCTATGATT CCCCTTGCTT GCGTTCTGT TCCACAGTTG AAAAATAATT CTCACTATCT	3600
GAAACATCCG TTATGATACC GGTGTGTCCA TAGATTCCAG GAGACAATTG GCTTTACCAG	3660
TTGGCCATGA AAGGATTCTT TTAAACGACA CTGAAATAAA AGAAAAGAAG CCAAATAATT	3720
ATAAGGCTTC CTTCTTGAA ATTGTTTTCG TGTCATTGAA AAAGCAAATA ACCACGGACA	3780
AAAAATTTGT CGAGTAATCG TATGTTCCAT ATCAAATGGA AAAATTCGAA AAATCCCCC	3840
CAAAAAAGCC AATCCTAAAC CATACCACAA ATGTTTGTGT TACGGGACGA GGTAACGCGA	3900
CCCCGTCCCA CAATTTATAA AACACTCGCT CTATATACCA AATCTTTGTC CAATTGCGTA	3960
TATTGCTCAT TTTATCGCCC CCATGACCGC ATGCGATCCA GAATCCAATT GAATACTCCA	4020
GACGTATCTC CAACTAACCA AACAGCTACT GCCAGCACAA GAATCAATGG AATGGCTTTG	4080
CTTGCTTTTC GTTCAAAGGC AATATAGGCC GCACCACCCC AAGCATCGCC AATATAATCA	4140
AATAATTAATA ATCCCCAACA AAATTGGTAA ACACATTTTG AAAAAATCA TTTGCTTTTT	4200
TCCTCCCTGT TAATACACAT TGAACATAATG ACAATACACA TCACCACAGA GCACAACTGA	4260
TAAAAATTC ACTTGATATT CCTCATTATG CATTGTCTT GTTCTTTGAT TGCTGTTTTT	4320
ACGTAAGAAT TATAAATAAT TTCTCACAG GCGATACATT ATTTTCACAG GAACATGCGT	4380
ACAATAAATC CGGTTCTGCG TACTTTTCTT TGTTTTAGGA TCAAGTTTCA TAGAATAGAA	4440
ATCATAAAAA TATTTACCAC TTTTTATCA TATGTTTCTT CGATCACTAG TGAAAAGCTA	4500
GGGTGGTAAA TCCAACAAAT ATAATTGATA CTTCCCTCTA CTGGATAAAG GGACAAAACC	4560
AAGTACCTTT AAATGGTCTG ATACAGTCAT AAACACGATC TATCTCCTCC CAAATAAAAC	4620
AAAAAAGGC TGTAATCGCA TTTTGATTAC AGTCTTTTTT TGTAATGTA TTAATTAAAG	4680
CTTTACTTCC GGTCTTTTAA TTTAATGACT TTAGTCAGTT GAAATCTCTA AAGAGATTTT	4740
GACAAAAAAA CCACGCGTAA TGC GCGTGGA GGTCAGAGCT TTAGCGATAA AAGTTTAGCA	4800
CATACTCGTT GGAATTGTAA GTATCATCTT GTGTTTACAC CAAAATATAG AAGAAAAGTA	4860
ATCTATGGTC AATTGAGGCG AGATATAGGG AAAATACTAA GAAAATTGTG TGAGATGAAA	4920
GAAGTCGAAA TCATTGAAGC GCATGCAATG CCCGACCATA TCCATATGTT GGTAAGAATT	4980
CCACCGAAGT TGAGTGTATC AGGTTTTATG GGTTCCTTAA AAGGAAGAAG TGCAGTAATC	5040
ATTCATGAAC GTCATGCGAA TTTAAATAC AATTATGGAA ATCGATCCTT TTGGTCAAAA	5100
GGTTATTATG TGAGCACTGT AGGACTGAAC CAAAAACGA TAGCAAAGTA TATACGAGAA	5160
CAAGAAGCAG AAGATCGAGT AAGGGATAGT ATAAACAAAA GAGAGTATAA AGATCCATTT	5220
AGGAAGTAAC AGCTAGATAA GAATAGAAAG GTAGCGGTTG GCGGTCTCTA AAAAAGCTCC	5280
CTTAAAGGGA GCAGCAGGTA GCGAGCCCTT CTAGGGCTAT TAAAAAGCCA CCCGCTATGC	5340
GGGTGGATAT TTACTTAATA AATCAATTTT CCCAAATTAT TTTCTTTAGC TTTCTTAACG	5400
ATTTCTGTTG CAACTACCAC ATCTAATACA GCCGAACCAA CTGTTTTGAA GATCGTGATT	5460
TGTTCTCCAC TTGTCCGACC GACCAATTCT TCATTGATCA GTTGGCCTAG TTCTCCATGA	5520

TAGCTGTCTC TTTGAATATA TCCATCTTGC AGCGGTGAAA TGaAGTCCCC AGCTTCTTTT	5580
AGAACACCGT CCATGGTATC AAAAATAACA ACATCTGCGG CTTTAATTAT TTCTCGTGGa	5640
ATTTACACA TTTCCGGTGT GTAAGCACCG ACTCCATTGA TGTGCGCACC TTTTTTGACC	5700
CACTCAGCAG AGAATGTCGC ACGTTTTGAA GTGGTCACAC TTGTGATGAT GTCTGCTCCT	5760
TCAACGCATT CTTGACTGGT TTTTGTTGGG CGCATCGTTA CAGAAAATTG TTGCATCATT	5820
TCTTCAGCAA ATTGGTTGGC ACGCTCAAAA TCAATATCAA AAATCCGAAC TTCTTCTAAT	5880
TTACGAACGG TCAACATTGC TTCTAATTGT GACTGTGCTT GTCCACCGGT TCCAATCAGC	5940
GCGCCGATTT TAGCATCTTT TTTTGCAAGA AGTTCTGTTG CTGCTCCTTG AACAGCGCCT	6000
GTTCTTAATT GTGTCAAATA AGTACCATCT AAACAAGCGG ATACCATAACC CGTTTCTGGA	6060
TCCAATACGA TCATCGTTGC TGGTACACTT GGTAAGTTTT TCTTGATGTT TTTAGGATAC	6120
ACGGAAACGA TTTTGACACC TAGTGCTCCT TCGCCACCTT CTACATAAGC TGGCATATAT	6180
AAGCTTTGTC CGTTACTTTT TGGCACATCG ATATTGGTGC GAAGAGGCAC AATTGCCTTT	6240
CCTGCTGAAT ACAATGACAA TGCTTTTTTG TCTGCATCGA TGGCTTCTCT CATTGAAAAA	6300
TTTTTGATAA TATCTTCTTT TTTAAGAAAT AGCATTTGAA TCGCTCCTTA ATTAGCTGCT	6360
GTAGCTTTTT TTAATTGACT GCTCAACATC GTATATGCAC TGTATAGTAA GACCAAAATC	6420
ACGATCCATT TCAGAAATATA GGTATCTAAA GATTTGACTA GGAATACAGC GACCAATACC	6480
CCTAATACAC CGAAGGTTGA AGTGAAAAGT GTGATTTTTC TGCTGTATTC ATCAAATTTA	6540
ATGAATTGAA CACTGCCGAT CGGTACTGAG AATGTACAAG CCCCCATCAT AATAGGAAAG	6600
GCTGCAACTG GATTCAAGCC TAGCAAGTAG ACAGTTACCA TTGTTAACGC ATAAGAACCA	6660
ATGCCGATAT GTTCAATGC GCCATAGACA AAGAGTAAAA ATCCTGCCAA GATTAATTTT	6720
CCGCCATACA ATTCACTGGC TGTTCGTTT GAGgGAAkCC AATTCATTG TCCCATAAAA	6780
ATTAAGAATG CAGCAATGAC CAATCCTATG CCGACAAACA GTTTGATCGT TTTCTCTGGT	6840
AACTTCACGA CAAATCTTGG TCCTAAATAT GCACCGATAA TTTGGCAGAT AATACAAACA	6900
GCTAATGTCT TGATCCCTAC GTTGATAGCA GTGATATAAG ACAATGCCAT GACGGCTACC	6960
GGTACTACAC ATTGGGTATT CAACGTTCCG GGTAGTTTTT TCATGCTGAC CCATTTCAAC	7020
TTTGGATAAA GGACGGTTC GATAGCAAAG TCGGAAACGC CAAATGTTGA AAGAAAGAAG	7080
GTGATAAATG ACGTGAAGGG CAACAGACGC ATATCCCCAG GTTCTTCTTT GAAATTTTGT	7140
TTGTTTTTGA AAAGGTCTCT TAGAAAGTAA ATAATAAAAT ATAGATTGAT TGCGACAATT	7200
AGTGCCAGCA AAATATTCCG AATCATAAAA ATTCTCCTA CTAGTAAATC TTTCTGATTC	7260
CCTCAAAGTT CTCACCATAT TTTGCTTTAA GTTCTTCAGA AGATTCTGCA ATCACTTTTA	7320
GTGTATATGC ATCATGTTTG ACATATTCTT TAGCTAGATT GACCATGGCG TGGTTTTTGT	7380
TCCAAGTGAC ATCGGTTGAT TGCAGTTCTT CATTAAATGGT GCCAAACACG ACTTCTTGCT	7440
CATCGCGCAG GATGTTGATC CATCTTGCCC CAAAGTCATC AAGTTTGTCT GTTTCAAAC	7500

CATGAATGTA ATACTTTTCT AAGGGAAGAT CATATTCATG GGTAGAACTG AATAAAATCA	7560
GTACAAATTT ATCGACTTTT TTTTCGGCCT GCTGCAATGC CTTTAGCAGT GAATCCgTGA	7620
GGTTCTCATG CCAAACCTGG ATGAGCAAAC TTTCTTTTGC ATTTTtCACC AAATGCTCTG	7680
CTTaTCCaGA ACATACTcGA TGCCATCAAC TTTnCCACAG CATGTCTTCT TCATCTTTTT	7740
CCTTAATCAT TcCTAGATTC TTTGTAAGaA AAGaAAGATC aTTTTTAACA CTTTTTTCTA	7800
ATTTCTCTAA AAACtCaTTT GCCGAGATGG CATGATAAAG TTTAGGTCTT GATTtATTGA	7860
CGACAATCAA ATTCTTTTTT AATAGTTTTT CCAAAATGTT ATATACTTTC GAGCGTGGCA	7920
CcCTGAGATT TtACTTACTT CATATCCTGT AAGATTACCT TTTTCTAATA GGGTGGTATA	7980
CACTCTTGTT TCCATCTCTG AAAAATCATA TTTTTTCATG ATTTGAATCA TCGTTTCCAT	8040
TGATTAATCC TCACTTCATT TTTAGTAGCT ACTACGGTAA CAACTATAAT TTTTTTTTGA	8100
TCATCTGTCA AGAGTTAGTT CAATAAAACA CATATAAGAA TTAGCTTAAA AACAAGCTTT	8160
tTATTAAaA TaAATaATAT AATAGGGA	8188

## (2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3467 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TGATGAGCAT AAGGAGGACA GGCATGAACA AAAAGAATAA TGGCATGAAA AATGGTTTAT	60
ACTACGTCTT AGTAGTACTT GCCATGGTCA GnGTTGTCTA TTTCATTTTT GGAAATAACA	120
ACCAACAATC GCCAGACATC GATTACTCAA CCTTCCAACA ACAATTGGAA GACGGAAAAG	180
TCAAAGACAT GACCATTCAA CCAACTAATG GGGTTTACCG GATTGAAGGT CAGTATAAAG	240
AAAAACAAGA AGTGAAAGAC ACAGGTGGTC TGTCATTGTG GGGCTCAACG CAAGCGTCGT	300
CAAAAGGCTT CACAACAACC GTTTACCTA GTGATACAAC CTTAGCTGGC ATTCAAGATG	360
CGGCACAAAA TAACAAGGTG AAGCTAGTTG TCAAAGAACA ATCAACAAGT GGTGCTTGGT	420
TGTCACTGTT GTTTAGTTTC TTACCATTAG TGATTATTTT CTTCTTCTTC TACATGATGA	480
TGAGCCAACA AGGCGGTGGC GGTGGCGGCG GTGGCCGTGT GATGAACTTC GGTAAATCAA	540
AAGCCAAAGA AGCCGATAAA AAAGCTAACC GTGTCCGCTT CTCAGATGTA GCTGGTGCGG	600
AAGAAGAAAA ACAAGAATTA GTTGAAGTAG TGGAATCTT AAAAGATCCA CGTCGTTTCG	660
CTGAATTAGG CGCACGGATT CCTGCGGGCG TACTATTAGA GGGACCTCCA GGGACTGGTA	720
AAACGTTACT TGCTAAAGCC GTAGCAGGTG AAGCAGGTGT GCCATTTTAC TCAATCTCAG	780
GTTCAGATTT CGTTGAAATG TTTGTGCGGG TCGGTGCTAG CCGTGTCCGG GATTtATTTG	840
AAACAGCGAA GAAAAACGCG CCTGCAATTA TCTTTATCGA TGAAATCGAT GCAGTTGGTC	900
GTCAACGTGG TGCTGGTATG GGTGGCGGAC ACGACGAACG TGAACAAACC CTTAACCAAT	960

TATTAGTTGA GATGGATGGT TTTGATGGCA ATGAAGGAGT TATCGTAATT GCTGCAACGA	1020
ACCGTTCAGA CGTGTTAGAC CCAGCGTTGT TACGTCCAGG TCGTTTTGAC CGCCAAATCT	1080
TGGTAGGACG TCCAGATGTC AAAGGTCGTG AAGsgATTTT aCGTGTTTCAT GCGAAAAACA	1140
AACCGTTAGC TGATGACGTT GATTTGAAAG TTGTTGCACA ACAAACACCA GGTTTTGCTG	1200
GTGCTGATTT AGAAAATETC TTAAACGAAG CAGCCTTAGT TGCCGCTCGT CGTAACAAGA	1260
AAAAAATTGA TGCTTCTGAT GTGGATGAAG CCGAAGACCG CGTAATTGCT GGTCCAGCTA	1320
AAAAAgATCG CGTGaTTAAT AAAAAAGAAC GCGAAATGGT GGCTTACCAT GaAGCAGGAC	1380
ACACAATTGT TGGTTTAGTC TTAAGCCGTG CTCGTGTTGT CCaCAAAGTA ACAATTATCC	1440
CTCGTGGTCG CGCAGGCGGT TATATGATTG CTTTACCAA AGAaGATCAA TTCTTAATGA	1500
CTAAAGAAGA TATGTTTGAA CAAATCGTTG GCTTACTTGG TGGACGTACA GCAGAAGAAA	1560
TTATCTTTGG CGTTCAATCA ACAGGGGCTT CAAATGACTT TGAACAAGCA ACAGGGATTG	1620
CTCGTAGCAT GGTAAC TGAA TATGGAATGA GCGACAAATT AGGACCTGTT CAATATGAAG	1680
GAAACCATCA AGTCTTTGTC GGTCGTGATT ATGGTCAAAC AAAAGCTTAC TCAGAACAAG	1740
TGGCGTTTGA AATCGATCAG GAAGTTCGTC GTATCTTAAT GGATGCTCAT ACGAAAGCGC	1800
ATGAAATTAT CGAAGCGCAC CGTGAACAAC ATAACTAAT CGCTGAAAAA CTATTAGAAT	1860
ATGAAACATT AGATGCAAAA GCAATCAAAT CATTATTTGA AACAGGTAAA ATGCCTGAAG	1920
GTGCTGACAG CGACTATCCT AGTGAAAAAG AAGCGCAAAC ATTCGAAGAA GCAAAACGTG	1980
CTTTAGAAGA AAAAGAGGCG CAAAAACAAG TTGAAGAAAA ACAAGACTTT GAAGAAGCGA	2040
AAAAAGAATT ACATGATGAA GCAGAAGAAG TCAAAGTAGA AAGCGAACAA ACAGAAAAAG	2100
AAGTCCAATC TGAAGAGAAa AAAGATTGAG ATTCCAATTC TGAATATGAT CGCAATAACT	2160
ACGAAGATCG TTATAAATAA CCAATTAGTA GAAGGTGCAG TATTTGTCGA GTTGACGAAT	2220
ACTGCACCTT CTATAATATT AAGATATTCT ATAGAAGGAA GATTATTTGG TTGAAAAAAC	2280
ACCTGATTCC TAGTATTATC yGACTAATGG aCmAAAAAGC ACCAAGAGAG nGAGGAAAAA	2340
GAATGGAAGA TTATTTAGTA AAAGCCCTTT GTTATAAAGG TTCAATACGA GCATATGCCA	2400
TCAGTGCCAC TGAAACTGTT AGCGAAcGCa aAGAAGACAT GATACATGGA GTTCTTCTAC	2460
TGCAGCGTTA GGACGCACAT TGATTGGAGC CCTTTTACTA GGAGCGACTT TAAAAGGTGA	2520
CGATAAATTA ACTGTAAAAG TACAAGGAAA TGGCCCAGCT GGCCTATTA TTGTCGATAG	2580
TAATGGCCGC GGAGAAACAA AAGGCTACAT TAAAAATCCA CATGTCAGCT TAAAGTTGAA	2640
TGCAACAGGT AAAATTGATG TACGTGGTGC TGTCGGTAAC GAAGGAATTT TCACAGTCAT	2700
TAAAGATTTA GGATTAAAAG AAACCTTTTC TGGACAAACA CCTATTGTTT CTGGTGAAAT	2760
CGGTGAAGAC TTTACATATT TTATGGCGGT CTCTGAACAA GTTCCTTCAG CTATCGGTTT	2820
AGGTGTATTG GTGGACACGG ATGAATCTGT GAAAGCAGCC GGTGGTTTTA TGATTCAAGT	2880
AATGCCAGGT GCCGATGAAA GTACCATTGA TTTTATTGAA CAGCGTCTAG CGGAAGTGCC	2940



ACCAATTTCA CAATTACTAG AAAATGGAGA AACACCTGAA CAA <sup>1</sup> GTCTTGT ATCGCTTATT	3000
GGGAG <sup>2</sup> AAGAC GAAGTAGAAA TTTTAGAAAA AATGCCTGTA CAATTTAAAT GTGACTGTTC	3060
CAAAGAAAAA TTTGCGACCG CTTTGATTGC TGTGGGATC GATGAATTAA ATGCGATGAT	3120
TGATGAAGAT CATGGAGCAG AAGCAGTTTG TCAATTCTGT AACAATAAAT ATCACTATAG	3180
TGAAGAAGAA TTAATTGAAC TTCGCGATGA AGCGATTGCG AATACGAAAC AAAAGTAGGA	3240
GGAAAGACTA TGTGGAAAAT TGGAGACGTT GAAATCCCCA ACCGTGTTGT CGTAGCTCCG	3300
ATGGCGGGTA TTAGTAACGC CGCTTTTCGT GTGACTGTAA AAGAATTTGG CGCaGGATTA	3360
GTCGTTTgTG AAATGATTAG TGACAAAGGa ATTAAATTAA GAAACAAAAA AACCTTAGAA	3420
tGCTATATAT TGATGAGCGG GAATAnCCTT TAAGTGACA ATTTTGG	3467

## (2) INFORMATION FOR SEQ ID NO: 23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21126 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CTTGAnACAA TTGGGGCAAA AAATGTTTTG GCAATGAAA AAGGCTGGTT ACCAGTGACA	60
GAAGAAGCGG CCATTGCTGC AAAACCAGAG GTCATTTTAA CCAATGTAA CTATATGAAA	120
GATCCCGCCA AAGAAATTTT AGCTCGTAAA AACTGGGAAA GCGTTCCAGC AGTTCAAAC	180
AAAGAAGTTT TCGAAATCGA TAACATGTCT AGCTCTTTAC CAAATAACCA CATTACCAA	240
GCATTAAAC AAATGGCAAA AGCTGTTTAT CCAGAGGAAT ACAAGGATTT AAAAGATGAG	300
TAAAAAACG TGTCTCATAC TGATGCTCTT GATTAGTATG AGCATGCTCT TTTTGGGTAT	360
TAA <sup>1</sup> AATTGGA AGTGTTGAGA TTTCGTTAAC GGATTGCTT CAAATTTTAG TGAAAAAGAC	420
GGTTTCAGAA GATGGGATTT TAGAAGGAAT CATCTGGAAT GTTCGGTTAC CGCGTGTCGT	480
GATGGCCTAT CTCGTGGGCG CAGGCTTAGC AGTAAGTGGC ACGGTGATGC AATCGCTCTT	540
AGGTAATCCT -TTAGCTTCTT CGTATACTTT GGGCGTATCA GCGGGAGCAT CCTTAGGTGC	600
TGCGTTAATC ATGGTAACTG GAGTAACTGC TTGATCTTA GGGGCATTTT TATtGCCTTT	660
AACTGGTTTT ATCTTTGGCT TAGCAACGGT CTTTTTAGTT TTGTTATTTA CGCAAACAAT	720
GGATAGTCAA ATGTCCAATC AAACAGTTGT TTAGTAGGA ATGATTATGA CCTTGTTTGT	780
CGGTGCTATT TTGACGTGA TTACAGCACT TTTCCAAGAT TATTTGAAAC AATTAGTCTT	840
TTGGCAAATG GGTAGTTTTT CTGGCAGTAA TTGGCAAAAA ATTGCTATTT ATTGTCCAAT	900
CTTGCTTGTC AGCAGTCTGT TTCTTTGGTT TGATGCCAAC GCATTGGATG TTCTCGGTTT	960
AGGAGAAGAA CATGCGATGT TAGCGGGAGT AGAAGTCAA ACAGCAAAAC TACGCATTAT	1020
TCTTTTGGCT AGTTTATTGG CAGGAAGTGC GGTTCATTT GTCGGTGTA TTGGTTTTGT	1080

TGATTTAATA GCACCACATA TTGTCCGTCG CTATTTAGGG GCGACTCATC GCTGGCTGAT	1140
TCCCGGTTCA GCAATCTTAG GTGGGACCAT CATGGTACTA GGGGACACGA TTGCCCCGAC	1200
GATTTTATCA CCAAGAGAAA TTCCCATCGG GGCCGTAACA GCCTTGATTG GAGCGCCATT	1260
CTTCTTGAT ATTTACTTTA AAAAAAGGGG CGTTGCTGCA TGATTGAATT AAAAAAACA	1320
GCTGTAACGC TCCAACAAAC CCCCATTTTA AAAGATATTT CCCTCGTTTT TCCGACAGGT	1380
AGTAAACTT GTATTTTAGG TCCAAATGGT TCGGGTAAAA CGACACTTTT AAAAACGATT	1440
GCCGGACTTG TCCCTTATAG TGGTCTGTT TTAATTGATG GTCAAGAAGT TCATGGACAA	1500
AAGCGGAAGG AATTGGCAAC GAAAGTAGCG ATGATGAGTC AATTCACGAC AGTTGCTTTT	1560
GATTATACGG TTTACGAAAC AGTTTGTATG GGGGCGTATC GACAACAAGC ACAACGTTTT	1620
TTACCTATTG TTTCTAAmCA rGAAAArGAA CGAGTCCTTw ATwATTtAGA ACGCACGGGT	1680
TTATTGCCTT TAAAAGATAA mGTTGTCAAT CAGCTTTCAG GAGGACAACA ACAACGAGTA	1740
TTTTTTGGCTA AATTATTTGT ACAAGATCCA GAAATTATTT TATTGGATGA ACCCmATAAT	1800
CATTTGGaTA TTCGGTATCA ACAAGAATTG ATTCAGCAGT TAAATGAGTG GAGTGCACAA	1860
GAAGGGAAAA CGTTAATCGG GGTGTTTCAT GATATCCGTT TAGCGTTAAC GCTAAGTGAA	1920
AAAATTGTGT TTATGAAACA AGGGAAAGTG GCCGCGCAAG GTGATTTTCA AACGTTGGCT	1980
TCCAAAGAAT TTTTACAAAC GATTTTGTAA ACAGATATTG TCTCCTATTT TCAGAAACAA	2040
CATAAAGTTT GGGAACTAT TCAATAATTT TGACAAGTTT TTTGGAAACA GGTAAGATAA	2100
AGGCAAAGAA TCAGAAACAC AATGGCGTGT TTTAACAAAT CGGCAAAGGA GCCCAAGACT	2160
AAGTACGTGA TGGTACTTGC TTTGGGCTCC TTTTTTATGG AGGTAACGAT GAAACGAATC	2220
ATTTTAATGG GAGCTATCGG TTGTGGCAAA ACCACGTTAT GTCAAGCTCT GCAAGGCAAA	2280
GAATTAATTT ATGATAAAAC ACAGGCGGTG GAATTTTCATA CAGAAATGAT TGATACACCA	2340
GGCGAGTTTA TTTTGCATCG ACAATATTAT AATGCATTAA ACGTCACAGC GGCAGAAGCA	2400
GATGTGATTG GtTTGGTTCA AAGTGTGTA GAAACACAAC AAGTTTTTTC ACCAGGATTT	2460
GGTAGTATTT TCCCGAAAGA AATCATTGGT ATTTTAACCA AGATTGATTT AGCACAAGAT	2520
TCTCAGCAGT TAGAGATTGT TCGGCAACAA TTA AAAAGTG CGGGTGCAaC AAGAATTTTT	2580
GAAATTTCTT CTGTTGAAAA AATTGGCTTG CAAGAATTAG TAGATTATTT AGAGGAGGAT	2640
GAAGCAGAAT GAAAATTTAT ACAAAAAC TGCGATAAGGG AATGACGAAA TTAGTCGGCA	2700
GCTCAACCGT CGCTAAGGAT TCTGATCGAG TTGAAAGTTA TGGAACGATT GaTGAATTGA	2760
ATTCTTGGGT GGGCTATATT ATCAGCCAAC TTCCACAAGA AAATCAAGAA ATTAAAGAAG	2820
AACTAGAAGC TTTACAACAT TTA CTTTTTG ATGCGGGAAC AGATTGTCA ACACCGATTG	2880
AAGCACAACG ACCATTTAAA TTACAAAAGG CAAGTGTGCA TTGGCTAGAA CAGCGGATTG	2940
ATTTCTACAC TGCACAATCA CCAGATATTG ATCGTTTTAT TTTACCTGGC GGAACCCAG	3000
CAGCCAGTAT GGTTCATGTG GCTCGGACGA TTGCACGGCG CGCAGAACGC ATCATTGTTC	3060

GTTTAAATTG GACAGCCAAA ATTAACGAAG AGGTATTAAT TTTTACCAAT CGTTTATCAG	3120
ATTATTTTTTA TGCGTTGGCT CGCTGTTTAA ATGTTCAAGC ACAACGACCT GATGTTTTTT	3180
ATGAACGAAG CGAAATGGTT TTCCATAAAA TTAAAGAAGA TGGATTGTAA ACAGTTGCTT	3240
TTCTAAAAAA CTTATGGTAA ATTTCAGATA GaTTAAaCGA ATATTTTATG AaAAATATGA	3300
ATGGAAGCCA GTGAGAATCT GGCACGGTCC CGCCACTGTG AAGAAGCAAG GTTGCTTTTA	3360
AGTCAGGTCT TTTCATTTTT TCATTATTGG GCATGCTGTT TCGAGGCAAA ACAGGATGTT	3420
TCTTAACAAC GCTTGTG!TG AAAATCCAGC CCAAGATATT TGTATTAATC CAATTAATGG	3480
CACGAGGAAG TGCTTTTGGG AATTTTGGGC TGGTCTTCA ATGTTGAAGG AGATTTTCAGG	3540
TACAATGACG TACTGCTATA TTCCCTTTTT TTATTGACAA TTAATTAAAG GCGTTGTACT	3600
CTATGGATAA GGTTCGTGT ACAATGGCGT ATACATAAGG AAGCAAAGAC GCTTCAGACA	3660
GATGATTTTT CGTTTGTGTTG AAGTGTCTTT TTGTATTTTA TTAAGGAGGA ATTGAGATGA	3720
AAACAATTCA TTTCCCTACC GAACTATGGG TAGGAGAAGG CGCATTAGCC AATTTGGAAA	3780
CACTTCATGA TAGACGTGTT TTCATTGTCA CAGATCCATT CATGGTTGAC TCTGGATTTG	3840
TGAATGAAGT GACAAAGCAT TTAACAAAA GTGAATGGCA AATTTTTAGT GATATCATCC	3900
CCGATCCACC AATTGATAAG ATTGCTGCGG GAATTAAGCA TTTAGCGACA TTTCAAGGAG	3960
ATACGATATT AGCTTTAGGT GGGGGATCCG CTATTGATGC GGCCAAAGCA ATGAAATTTT	4020
TTGGCAAACG TACTTTGCAA ACTCAGATTG CTGAGTTTAT TGCGGTTCCCT ACAACTAGTG	4080
GAACTGGTTC AGAAGTTACG AATTTTTCAG TGATTACGGT TGCTGAAACT GGCACCAAAA	4140
TCCCGCTGTT AACGGATGAA ATTCACCCAG AAATTGCGAT TTTAGATACA AATTTAGTGA	4200
TGAGTGTCCT ACCAAAAAIC ACAGCGGATA CAGGAATGGA CGTCTTAACG CATGTTATTG	4260
AAGCGTATGT TTCAACAGAA GCAAATCCGA TTTCAGATGC GTTGTGCGAA AAAGTCGTTT	4320
GCTTAGTATT CGACAATTTA GAAATAGCCT TTAATGAGGG TAGCAACCAA CAAGCCCGTG	4380
AAAATATGCA TTTAGCTTCA TGTATGGCAG GGATGGCGTT TAATGTTACT tCGTTAGGCT	4440
TGAATCACGG TATTGCTCAT GCGGCCGGTG CACGTTTACA TGTGCCACAT GGACGAATGA	4500
ATGCAATGCT CTTACCTGAA GTCATTGCAT ATAACAGTGG CTTAGCGAAT GGCAAAGTCA	4560
CCAACGAACC AACAGCTAAA CGTTATGCGC AGTTAGCAAA TTGTCTAAAT GATACACAGA	4620
CAACCAATGC ACGGATTGGT GTTCAACAAT TTATTCGACA AATCAAGCAA CTCCGTCAAA	4680
AATTAAACAT GCCTGCCACT TTTAGTGAAT ATGGTTTACC AAAAGAAGAA GTCCAGGCTG	4740
CAATCCCAAA AATAGCTGAA GCGCATTGA TGGATGGCTG CACCAAAACA AATCCTGTCC	4800
AACCAACAGC AGCAGAAGTG AAAAAAATTT TAAATTCTAT CTTATAAGCG ATTGACAGGA	4860
TGAAATCGAT ArCATATAaT GTGGGTAGCA CAACGGCGTG CTTcmAAATT TAAGAGCAAA	4920
GAAGCTCCTT AGTAGAAAGC GAAAGTGGA CTTTACGTT TTTTATTGAG GATTTTTTAT	4980
TTGCTAAATT AAAGGAGCGA GTCTGGTTGG AAGAAAAACA ACGAATGATT CAAGAATATG	5040

TTCTGGGAA ACAAGTAACG TTAGCGCACA TCATTGCTAG TCCTAATAAA GAAATTTATA	5100
CAAAATTAGG ATTGCCAGAA GGCACCAGCA ATGCACTAGG TATTTTGACA ATCACTCCGA	5160
GCGAagnCCA ATTATCGCTG TTGATATTGC GACTAAAAGC GGAGATATCC AAATCGGTTT	5220
TATTGATCGT TTTTCTGGAT CGGTAGTCAT TTCTGGTGAC GTAAGCTCAG TGGAAGCAGC	5280
CTTACAAGCG GTGATTGaAG GTTTACAACA GATTTTAAAC TTTTCTGTAA CTTGTAAAAT	5340
TACCAGAACA TAGCGTTGAA AGGACAAGGG CGTGAAGAAA ATGGATGGAC GAATTGTAAT	5400
AGTCGATGAT GAACCTATTA CGAGACTCGA TATTCGTGAT ATTGTGATAG AAGCAGGGTA	5460
CGAAGTAGTC GGTGAAGCAG CTGATGGTTT TGAAGCAATT GAAGTATGTA AAAAAACACA	5520
ACCTGACCTA GTATTAATGG ACATTCAAAT GCCAATTTTA GATGGTCTAA AAGCAGGTAA	5580
AAAAATTGTT CAAGACCAAC TAGCCAGTAG TATTGTCTTT TTATCTGCAT ACAGTGACGT	5640
ACAAAATACG GACAAGGCTA AAAAATTAGG CGCACTGGGT TATTTAGTTA AACCTCTAGA	5700
CGAAAAATCA TTAATACCTA CAATTGAAAT GAGCATTGAA CGAGGCAAAC AAACGCAGCT	5760
GTTACTAAGT CAAATCGATA AATTAAGTTT AAAATTAGAA GAACGTAAAA TTATCGAAAA	5820
AGCCAAAGGT ATTCTTGTA AAGAAAATCA TATATCGGAA GAAGAAGCCT ACCAAATGTT	5880
GCGTACGTTA AGTATGAACA AACGCGCACG TATGAGTGAA ATTGCAGAA T GATTGTAAT	5940
GGATGATGAA TAAATGAAAC GATTAGAGCA ATTATGTCAC CAATATACCA ATCTTTCAGA	6000
ATCAGACATT AAAGAATTAC AACGTACAGC ACGTTATCTT TCTTCGACGA CACTCTATCa	6060
AAGTGCAGAT GTTTTTATTG ATGTGTATAA GGAAATGTCA CAGCAAGCAC TAGTCGTTTA	6120
TCATAAGCCG CCTGCGAAGA CAACTTCTTT ATATAGTGGA GATGTCGTCG GGATGGAAgC	6180
TTTATTAAAA AATGAACCTG GTGTTCTGCG TACCATGCAA ACTAGCtTAA ATAGCATTGG	6240
TTtGTTAGCA GTGACTCAGG AAAATCGTTT GATTAGCAA AATATCTATC CCATTTCGAA	6300
TGAACATCGA ACGATTGGTG TCATTATTGT CGAGATTGCG GCAGATGAAG AGATTCAAGC	6360
GGACTTACAA AAAGAGGAGC TAAATAACTG CCAGTTAGCA AAAGTAGCGA AATCAACTAG	6420
TCAAGTCGAC GCATTATTTA TTGATCAATT GGCAGAAGCC GTTTTGATT TTGATGCGGC	6480
GGGTCACTTA TTAATTACCA ATCACAATGC ACAAGAGCTT TATCGCAAAC TTGGTTATCG	6540
GGATAATATC ATTGGAATGA GTTATGACAA TTTGTCCATC GATTATACAA CTTTGAATA	6600
TGTGTTGTAT CAAATGAAAT ATAAAATGAG TAATCAACCA ATTGAAAGCA AAACGACCTA	6660
CCTTAATTAT TATTTTAAGG TTCGGAAAGT GTGGTTGGCA TCAGAAGAAC AACTGATTAT	6720
GATTATTCAA GACAATACAG AATTTAAAGA AAAAGAAGCG GAAATTATTT CCAAATCTGT	6780
TGcGATTTCGT GAAATTCATC ATCGTGTTAA AAATAACTTG CAATCGGTGG kTTCCTTATT	6840
GCGTATTCAA GAACnGGCGA ACGCAAAGCC CCGAAGCAAA GAAAGTCCTT CATGAAAGTG	6900
TCAATCGAAT TATGGCGATT GCAGCCACCC ATGAATTGCT GTCAAAACAA GTCAAAGATG	6960
ACGTTGCTTT ACGGCAAACA TTGGAAGCAG TGATGTATAA TTTTAGACAT CTTTTTCAAG	7020

GCGCACAACC CATTGAGATG ATGATGGATG TTGATCCAGC CATTATGGTT TCTAGCGAAC	7080
AAATGGTTAC CATTTCCTT GTTGTCAACG AACTATTACA AAACATTTTT GATCATGCAT	7140
TTGAGCCACA AACAAGTGGC GTTGTAAAC TTAGTGGAAC ATTAGATAAT AAAATGATTA	7200
CCATTACTGT TACAGATAAT GGTAAAGGCT ATGATGTTCA TCAAAGTAAC GAGACCAGTT	7260
TAGGTTTGAT GATTGTCAA AGCTACGTAA AAGATAAACT AAAAGGAAAA ATTACAATTG	7320
AATCCAATA ACAGGGAAC AAAACCTGTT TCTATTTCTG ACAGAACACC AGTGATGTTG	7380
TTCATTGATT AAGCAAAGGC GCTTAAAGAA AAGTAGAAAG AAATTTTCTA ACGATTCTTT	7440
AAGTGCCTTT TTGTATGCAA AAGGAGGCGA GTCAGGCGAT GTCGAAGGAA ACACCTTTAA	7500
CCGTAGGAAT TGATTTAGGG ACGTCTACAA CCCAATTAGT CTTGTCAGAA TTAACGTGTG	7560
AAAATTTTGC CTCTGCGTT ACAGTACCAC GTATCAGCAT TTCTGATAAA AAAGTCATTT	7620
ATCGAAGTGA CATTATCTTT ACACCCTTAC TTAACCAATC AGAAATTGAT GCAGAACCTA	7680
TCAAAGCGTT TGTAGCGGAG CAATATCGCC AAGCAGGCAT TCATAAACAA GATATTCAA	7740
TGGGTGCGGT AATTATCACA GGTGAAACTG CTCGAAAAAG CAATGCCAAC AATGTGCTAC	7800
GAGCACTTAG TGGCTATGCA GGTGATTTTG TCGTTGCTAC GGCTGGTCCT GATTAGAAA	7860
GTATCATTGC CGGGAAGGC GCTGGTGCTC AACTTATTC TGAAACGAAA CGCAAACCAG	7920
TGGTTAATTT AGACATTGGT GGTGGAACGA CAAATTTAGC TGTCTTTAA GATGGCGAAg	7980
TGATTGACAC GGCTTGTTTT GACATTGGCG GACGCTTAAT TAAATTAGAT CAACAACAA	8040
AAATCACCTA TATAGCTCCG AAGATTCAAG AAATTATAAA TAAAAAGGG CTAACCTGC	8100
ACCTTGGTGA CCAAGCCACC GAACAAAAC TACTGCCAAT TATTCgGAA TTAGTTGCCG	8160
TGTTAGAAAA TAGCATTGGC TTGGGGACAC AAAGTCCGTT CTACCAGTTA TTAGTAACCA	8220
ACCATCCACT ACGAAAAGGT GAGGAATTAC CAATTGTAAC ATTTTCTGGT GGTGTGGCTG	8280
ATTGCTTAAA TACGACAAGT ACGAATCTCT TTAAATATGG GGACATCGGC TTACTTTTAG	8340
GGAAATACTT GAGAAAGTCA CTGATTTTTT CGGAAAAAGA AGTCTTAGAA AGCGCAGAAA	8400
CAATCCGAGC GACGGTCGTC GGAGCAGGTT CACACACTGC TGAAATTAGC GGAAGCACGA	8460
TCGCTTATCG TGAGCAAATT TTGCCGTAA AAAATATCCC AATTCtAAAA TTAGCGCAAG	8520
AGGATGAAAC CCTTACTGTG ACAGAATTAG GCCAGCGAAT CCAAGAAAA CTCAATTGGC	8580
ATCGTATTGA AGAAACACCC CAGATAGCCC TAGCAATTAG AGGCATGAGC AATCCAACAT	8640
TTGCTGATAT TCAACGCTAT GGTCAAGGGA TTGTGGAAGG ATTAGCTAGT TTAGTTGCTG	8700
AACAAATTCC CATTATTGTT ATGGTGGATG AAGACATGGC GAAAGCTTTA GGGCATGCCT	8760
TAAGTGCCCA TCTGCCAAAA GACTATCCGT TTATCTGTCT TGATTCCGTC AAAGTAGAAA	8820
ACGGTGATTA TGATGATATC GGTTTACCAG TGGCAGAAGG AGCCGTCCTT CCAGTCATTG	8880
TTAAGACCTT AGTATTTAAT TAAAAAGGAG CGACACAAAG ATGATTTTAA AAACCAAAC	8940
ATTTGGCAAA GTCTATCAAT TTACATCCGT TAAAGAAGTC TTAGCCAAAG CCAATGAAGA	9000

AAAGTCTGGC GACAAGTTAG CTGGCGTAgC TGCTAATTCA GCCGAAGAAC GCGTGGCAGC	9060
AAAAGTcGTT TTATCGGAAT TGTCCTTAAA TGACTTGTTT AACAAATCCCG TCGTTGrTTA	9120
TGACGAAGAC GAAGTGACAC GCATTATTAT TGACCAAGTC aATATGCGAA TTTTTGAAAG	9180
TATTAAACAT TGGACAGTTG CAGAATTACG CGAATTTATC TTATCTTCGG AAACaACCGA	9240
TTTTGATATC AAACGAATTT CCCgTGGACT CACTTCAGAA ATGATTGCTG CTGTCTGTAA	9300
ATTAATGTCC AACATGGATC TAATTGTTGG CGCAAAAAAA ATCAATATTG AAAAAACCGC	9360
CAACACAACA ATTGGCCGTC CCGGAACTTT TTCAAATCGT TTACAGCCAA ACCATCCAAC	9420
CGATAATGTC GATGGCAITA TGGCATCAGT TATGGAAGGT CTTTCTTATG GCGCCGGTGA	9480
TGCCTTAATT GGTTTGAATC CTGTTGATGA TTCTACAGAA AGTGTTAAAC GGATCTTAAA	9540
CAAATTTGAA GAGTTCCGCA GTGAATGGGA AATTCCAACA CAACTTGCG TGCTTGCTCA	9600
CGTAACGACA CAAATGGAAG CCATGCGTCA AGGCGCGCCA ACTGGCTTAG TCTTTCAATC	9660
CATCGCTGGT TCAGAAAAAG GAAACACTGC ATTTGGTTTA AATGCAGAAA TTTTAGCTGA	9720
AGCACAAGAT TTAGCGCTTC ACAGTGGTCA AGCAGCTGGT CCTAATGTGA TGTATTTTGA	9780
AACAGGTCAA GGCTCAGAAC TTTCTTCTGA AGCGAACTTT GGCGCAGACC AAGTCACAAT	9840
GGAAGCGCGT TGTTATGGCT TAGCCAAAAA ATTTGATCCG TATATTGTAA ATACCGTTGT	9900
CGGCTTTATT GGACCAGAAT ATCTTTACGA TTCAAACAA GTCATCCGCG CTGGTTTAGA	9960
AGATCATTTT ATGGGCAAAT TAACTGGAAT TyCAATGGGC TGTGATGTCT GCTACACCAA	10020
CCATATGAAA GCTGATCAAA ATGATATGGA AAACCTTAGCA ATGTTATTAG CAACAGCTGG	10080
TTGTACGTAT ATCATGGGGA TTCCTCATGG CGATGATGTC ATGTTAAATT ATCAAACAAC	10140
TGGGTTCCAT GAAACTGCGA CGATTCGTGA AACATTAGGC TTACGTCCAA TTAAAGAATT	10200
TGAAGAATGG ATGGAAAAAA TGGGCTTAAT GGAAACGGA AAATTAACAA GTCGCGCTGG	10260
CGATGCATCT GTATTTATTA AATAAACGAG GTGAGAATAC AAATGAACGA AAAAGAATTA	10320
AAAGAAATGA TTGCCGGAAT TTTAACAGAA ATGGTCGCTG ACAATCAAGC GGTTCGACT	10380
GCCACTGTGA CTGCTGAAGA AAAACCAGTC ACAACACACG TCACTGAAAC GACAGAAATT	10440
GAAGAAGGCC TAATTCCTGA CATTACTGAA GTAGATTAC GTAAACAATT ACTATTGAAA	10500
AATGCCGTAG ATCCTGAAGC ACTATTAAAA ATGAAAGCTT TCTCGCCAGC TCGTTTAGGC	10560
GTTGGTCGTG CAGGGACGCG TTATATGACT AGTTCAACAT TACGTTTCCG AGCAGACCAT	10620
GCGGCAGCAC AAGATGCAGT CTTTTCAGAC GTAAGCGAAG ACTTGGTAAA AGAAATGAAT	10680
TTTATTTCTA CAAAAACAAT CTGTAACAGT AAAGATGAAT ATTTAACACG TCCAGATTAT	10740
GGTCGTCAAT TTGATGAAGA AAATAGTGAA ATCATTCCGA AAAATACGAC ACCAAAAGCC	10800
AAAATCCAAA TGGTCGTAGG GGATGGCCTT AGTTCAGCTG CTATCGAAGC GAATATTAAA	10860
GAAGTCCTAC CaGCMATTAA ACmAGGATTA AACATGTACm ACTTaGATTT CGACmATGTA	10920
GkTTTTGtTA AAtATTGtCg CGTGcCAGCA ATGGaTaAAA TTGGTGAAAT CACTGACGCT	10980

GATGTTGTTT GCTTGCTAGT TGGCGAACGT CCTGGTTTAG TAACAGCTGA ATCTATGAGT	11040
GCCTACATTG CTTACAAACC AACTGTCGGT ATGCCAGAAG CACGGCGGAC AGTTATTTCA	11100
AATATTACACA AAGGTGGGAC ACCGGCTGTT GAAGCAGGGG CCTACATTGC CGAAATTATT	11160
AAAAAATGT TAGACAAGAA AAAATCTGGT ATTGATTAA AAGAAGCTGA ATAAAGGAGG	11220
ATTTCTCATG AAAAATGATC GTCTAGGCGC AAATGTCCTA AGTATGAAAG TCATTCCCAA	11280
TGTCGATCCG GCATTAGCAA AAGAATTATC ATTGAAACCA GAACATCGCT CATTAGGAAT	11340
TGTTACTTCC GATTGTGATG ATGTCACTTA TGTTCATTA GATGAAGCAA CCAAAGCCGC	11400
CGATGTTACT GTAGTTTACG GTAAAAGTAT GTATGCTGGC GCAgCAAATG CTTCAACCAA	11460
ATTAGCCGGT GAAGTGATTG GCATTATTGC TGGTCCGAGT CCTGCCGAAG TCAATAGTGG	11520
CTTGAGTGTC ATTACGCAAG TCATTGAAGA AGAAGCCAGC TTCTATAGTG CTAACGAAGA	11580
TGACAGCATT GTTTATTTTG CACATGTAGT ATCAAGAACG GGTAGCTTTT TATCTGAACA	11640
AGCGAATATT CCTGAAGGTG AAGCGATTGC TTATCTGATt GCACCACCGC TAGAAGCAAT	11700
GGTTGGCTTA GATGCAGCAA TGAAAGCTGC CGATGTTCAA ATGGGTGTCT TTTATGGCCC	11760
ACCTTCTGAA ACGAACTTTG GCGGCGCCTT ATTGACAGGT AGTCAATCTG CATGTAAAGC	11820
TGCTTGCACT GCGTTTGAAC AAGTCATTCA AAATATCGCT GACAATCCGT TGAGTTATTA	11880
AGGATGTGAC CTGAATGAAA GCATTAGGAA TGGTCGAAGT CAAAGGATTT TTAGGTGCCA	11940
TTAGTGTTGGC CGATGCGGCA TTGAAAGCCG CCGACGTTAC TTTGTTAAAA GCTGAAATTA	12000
TTAATGGCGG TTTAACAACCT GTGGAACCTGA TTGGTGACGT TGCCGCGGTT CAAGCTGCCg	12060
TTGAAGTAGG TACGGAAGTA GCTAAAGAAT TAAATTGTTT AATTGCTCAT CATGTGATCT	12120
CACGGGTGA TGCTCAAACG GAAGTGATTT TATCAGATCC AGAACCCAAG TCAGCACCTG	12180
AACCGATGGA ACAAGTAGAA GTAATTGAAG AAGAGATTGA AACACCAGAT TTGGAAGAAC	12240
ACACAGGCAC ACGCCAAAAA TTAGAAACAC AACGCGTGGT TGATTTACGC AAACAAGCCT	12300
ACAAAATGAA CCTAAGTAGT TTAAAGAAAA GTGAGATTAA ATTTGCCAAT AAAGAAGCTT	12360
TAATTCAAGC AATCATGGCT GAAATAGAAA GAAGTGAAGA CGAGTGAAT TAGTAGATAA	12420
AGACTTACGT TCCATCCAAG AAACAAGAAA TTTAATCAGA AAAGCCAAAG AGGCACAGCA	12480
AGTATTAGCA ACATTTTCGC AAAAACAAT CGATGCGATT GTTCAAGCGG TTCAGAGGC	12540
TACATTTAAC CAACGAGAAA AGTTAGCGAA AATGGCCAAT GAAGAAACAG GATTTGGTAT	12600
CTATGAAGAT AAAATCATCA AAAATGCTTT CGCTTCAAAA GTAGTATATG ACGAAATGAA	12660
GGACAAAGCT ACAGTTGGTG TGATCCATGA TGATGCTGCC AAAAAAGTCA CTGAAATTGC	12720
TGTCCCAGTG GGTGTAATTG CAGGATTAAT TCCTTCTACC AACCCGACAT CAACAGTCAT	12780
TTACAAAGCA TTAATTTTCAT TAAAAGCTGC CAACAGCATT GTTTTTTCAC CACATCCCAA	12840
TGCGTTGAAA TCGATTATTG AAACAGTGGA AATTATcCAA AAAGCGGCTA TTGCTGCAGG	12900
GGCACCAGAA GGTGTGTCT CGGTCATCAA AACGCCAACC ATGCAAGCGA CCAGCGAATT	12960

AATGAAAAAT	AAAGAAACAA	ATTTAATTTT	AGCTACTGGT	GGAAACGCAA	TGGTTAAAGC	13020
CGCGTACTCT	TCTGGAACAC	CAGCAATCGG	TGTTGGGCCCT	GGTAATGGAC	CAGCCTACAT	13080
CGAACGAAGC	GCAAATGTTC	CTCATGCAGT	GAAACAAATA	ATGGATTCAA	AAACTTTCGA	13140
TAACGGTACA	ATTTGTGCAT	CGGAACAATC	AATTATTGTT	GAAACTGTCA	ACCGTGAAGC	13200
GGTCaAAGAA	GAGTTAATTA	AGCAGGGAGC	GTATTTCTTA	AGTCCaGCCG	AAGCCGACAA	13260
ATTAGCTAAA	TTTATTTTAC	GACCAAACGG	TACAATGAAC	CCGCAAATCG	TGGGACGTTT	13320
CGTTCAACAT	ATTGCTTCAT	TAGTTGGTTT	ATCTATTCCA	AAAGATCGCC	GCTTAATTGT	13380
TGCCGAAGAA	ACAcATGTAG	GCTTGAAATA	TCCATTTTCA	AGnGrAAAAT	TGGCACCGaT	13440
TATTGCTTTC	TATACAGTTG	AAAAGTGGGA	AGcAsTTGCG	CGTTATCAAT	TGAAATTTTG	13500
AAAGGTGAAG	GTGCGGGTCA	TACAATGGGC	ATTCACACAG	AAAACAAAGA	AGTCATTGCT	13560
GAATTTGGCT	TAAGAAAACC	CGTCTCTCGT	TTGCTTGTTA	ACACCTCTGG	TACACTTGGC	13620
GGCATTGGTG	CTTCAACTAA	CTTAGTGCCA	GCATTAACAC	TTGGCTGTGG	CGCAGTTGGT	13680
GGCAGCTCAA	CTTCTGATAA	TATTGGGGTT	GAAAATCTGT	TCAATTTACG	TCGTGTTGCA	13740
TATGGCGTGC	GTGATTTAGA	AGAGATTGCT	CAAGAATTTG	GTCAAACATC	AACCACATCT	13800
GTGGCTACTT	CTTGCGAAAC	AACCAATCAA	GAAGAACTAG	TAAATGCCGT	GGTAGCTCAA	13860
GTTTTAGCTC	GCTTAAACTA	ACAAGAAACG	AGCGAGTAAG	AGTAAAACAA	ATAAATTATT	13920
GGATTTAGGG	TTTGGGAUCT	AAGTCAAATT	GCTAATTGTC	CCAAATTTTT	AAAAATAAAA	13980
CATACATCTT	AGGAGGAAAT	TATAATGTCA	AGTACAAATG	CTTTAGGAAT	GATCGAAACT	14040
CGTGGTTTAG	TAGGTGCTGT	TGAAGCAGCC	GATGCAATGG	TGAAAGCGGC	AAACGTTACT	14100
TTAATTGGAA	AAGAACAAGT	CGGTGGCGGT	TTAGTTACAG	TAATGGTTCG	CGGTGACGTT	14160
GGTGCTGTAA	AAGCAGCAAC	AGATGCTGGT	GCTGCTGCAG	CAGAACGTGT	TGGCGAATTA	14220
TTATCTGTCC	ATGTAATTCC	ACGTCCACAT	ACAGAAGTAG	ACGCTATTTT	ACCACACGCA	14280
ACAACTGAAT	TATAATTTTG	ATTTTGTGGG	GCAGAGGTAG	GTTTATTACT	TCTGTCCTAA	14340
ACTTTCATTG	AGGAGGGGTG	TTATGACAGT	TATTACAGAA	GATATGCTTC	GCATCAAAAG	14400
GCTAAAAAAA	GAAC TAGAAG	ATGGCACGGA	TTTTGTGTTT	CCTGTAGGCA	GTTTTCTGAC	14460
GCCAGCCGCC	AAAAGCTATC	TACGCGAGCA	TCGGATTAAT	AGCAGCTTCT	CTTCAACAAA	14520
CCCATTGCTT	TCTGAGGTTT	ATTCAGTGAA	GGCCTCCTCA	CAAAAACAAG	AACATTTATC	14580
GAAAGAAaCA	ATTTATGAGA	TTCGGCACTT	AAGCCATCTA	TTGTATTTGC	CTTTCTTAAC	14640
GAATGAAGCG	TTATCAACAG	AAGGTTGGTT	ATATTTCGAA	CAACAACAAC	AGTGTTTAGA	14700
GAAATTTATT	TTAGAACAGT	CGCTTGTAAT	GGAACCAGTC	ATTTTGGAAT	CAAAGGTAAC	14760
TATTCTAACT	AACCAACAGA	GACAATGGCG	CTATTCCTCA	AAAGAAATTC	AATTACAAC	14820
AGATAAAATT	ATTCTCCAGT	TAAAAGGAGA	ATCACAAC	TTCGGACTTT	TTCAAGCTTG	14880
GGCTACAGAA	TTAATTAGCA	CCATGGCTGA	AATATCACGA	AACTAGGAGA	TGAACAAATT	14940



GGATTCAATG AATGAAATTG TTGAAGAAGT TGTTAAACGA ATTcAGCAAC AACAAACAAA	15000
TACTTTTGAA GTTGAAGCTT CTGGCCGACA TGTTcATTtG TCTCGTCaAG AGATTGATGC	15060
ATTATTTGGA CCAGGATATC AACTAACAAA AGTGAAAGAT TTATCACAAC CTGGACAGTT	15120
CGTTTGTAAG GAACGAATTA CGGTTGCTGG ACCTAAAGGA CTCTTTCAAA ATGTGGTTAT	15180
TTTAGGCCCT GAACGTAGCG AATCACAAGT CGAAGTTTCT ATGACGGATA CGCGTATTTT	15240
AGGAATTAAT GCCCCAGTTA GAGAGAGTGG CAAGACAGAA GGGACACCTG GCGTCACGCT	15300
GATGAACGGA TCAGCAGTTG TGACATTAAG TCACGGCTTA ATTGTTGCCA AACGTCATAT	15360
TCACATGACA CCCGAAGACG CCTTGAAAAA TAAAGTAAGC AATAGCCAAA TTGTGCAAGT	15420
CAAAGTGGA GGCACACGTC CGTTAATCTT TGATGATGTG GTTGTGCGAA TCAGTCCGCG	15480
ATTTGCAACG TACATGCACA TTGACTACGA TGAAGCCAAC GCTTGCGGCT TAACCAAAGG	15540
AGCTCGTGGC TACATTCTTA AGTAAAAGGA GGCAAGACAA TGaACGTTTC TGrAATgGAA	15600
AATTTGATTC AACAGATTAC TGATCAAATT TGTCAACAAT TGCTGTCGTC GTCTTATTCA	15660
GTGGCTTTAG CAGAAGCAAA AGAAGCATTT CCTGAAGAAA TGTTTGTTTCG TTTTCCTGAT	15720
GTAAATTGGG TAACCAAGGA GCAGAAGCTA GCACGAGGCT TGGTTGTAAA GAGGCTAACG	15780
ATTTCCCAAG TAAATGCGAT TGCGCATTTA CAGGAAACGG ATGAATTAGT CAAAAACATC	15840
TTAGCATTCT TATTTGAAGG TAAGCCgGTA TTGGTTTTAA CCCCATAACC TTCAGTGAGT	15900
AAAAATTAC GATTGAAGTA TCGTTTGAAA CAAACGATTc AGGAAAACGT GGACATGTGC	15960
CAACAGTTTG GGCTGATTTT TTATCATGAT TCAGAAAATT ACGCTGTGTT TCAGGCGGCC	16020
TGTCAAAAAC AAtTCGTTCT TTGGCTGAGA CCAACGAAC CTACATTACT GAAAAACAAT	16080
TGATAAGAAT GACAGAATCA GGTGTTTCTT TAAGTAAAAA CGCCTATTTA ACCCCGTTAG	16140
CGAAAGATTA TGCTAGGAAA CATCAATTGC TGACGTAAAG AGGAAAAACG ATGCTTTTAG	16200
GAAAAGTGAC TGGTAGCCTT TGGTCAACAA GAAAAGATGA AAAATTAAAT GGCTGGAAAT	16260
TTATGATGGT TGATATTTTG AATCAAGAGG ACGAAAAGCA AGGATTTTTA ATTGCTGCTG	16320
ACAACGCTGG AGCTGGTGTC GGGGACAAAG TGTTAATTAG TCAAGGCAAT GCTGCCCCGT	16380
TTTCGGCAGA AGATCCGAAT GTCCCAATTG ATGCAATGAT TGTTGGTGTC ATTGATTCTA	16440
CAGCGGAAGA ATAGGGAGGA CGAGACATGA GCATTAAATGA AATTATTATG TATATTATGG	16500
TTCTTTTCAT GATTATTGGA GGAATCGATA AATGTATTGG CAACAACTA GGTTTAGGGG	16560
AGCAATTTGA GGAAGGGATT ATGGCTATGG GTTCTCTAGC GCTTTCTATG GTAGGGATTA	16620
TCACTTTAGC GCCTGTCTTA GCCAATCTTT TAAGTCCGAT TGTCGTTTCT GTCTATGAAT	16680
TACTAGGGGC TGATCCAGCA ATGTTTGCCA CAACCTTATT GGCCAATGAT ATGGGCGGCT	16740
TTGCGTTAGC GCAACAATTA GCCAATGATC CGCAAGCTGG TTTATTTGCG GGGGCAATCT	16800
TAGGTGCAAT GATGGGACCA ACGTTAGTCT TTACAATTCC AGTAGCGCTT GGCATyATCC	16860
AAAAAGACGA CCAACAATTC TTAGCAACAG GtGTTTTgAG CGGGaTTATT ACaATTCCTT	16920

TTGGTTTATT AGCAGGTGGT TTAACAGCTG GCATGCCATT GTCATTGATT ATTCCTAACT	16980
WAATTCCAAT TATTATTGTT GCCGCTTTAA TTATTtTAGG TTTATGGTTA GCGCCAAAGG	17040
GCATGATTAA AGGCTTCCAA ATTTTGGTC AAGGCGTAGT GATTGTCGCT ATCTTTGGTy	17100
TAGTTGTGGG GCGGATTCAA CTATTATTAG GAATTACAGT GATTCCTGGC ATTGCACCAG	17160
TTACTGAAGG AATTGAAGTC GTTGGCGGGA TTGCTTTAAC TTTAGCTGGC GCTTTCTGTT	17220
TAGTAGCTGT GATTACGAAA GTCTTTAATA AACCATTGAT GAAACTAGGA AAAGTTTTAG	17280
GAATGAATGA AGTAGCGGCT GCAGGAATGG TTGCTACATT AGCTAACAAT ATTCCTATGk	17340
kCCAAATGTT AAAAGACATG GATAATCGAG GGAAAATTAT TAATATCGCT TTTGCTGtTt	17400
CGGCTTCTTT CGTTTTAGGT GACCACCTTG GTTTTACTGC AGGCGTACAA AAGAAATGAT	17460
TTTTCCAATG ATTGTTGGCA AATTAGTTGG AGGGATTACC GCCATTTTTG TGGCTATCTT	17520
TATGGCAAAT CGAATGCTAG GGAAACCCGA AACAGAAGTG AAATAATGAG GTGTAAATAT	17580
GACAGAATTA AATCGTGAAA TGATTGAAAC GCTTGTTCTGA CAAATTGTCA CAGAAAAATT	17640
AATGCCGACT AAACAAGTGG ATCCAAGTGG AATTATGTCC ATCAAATTAC CCGAAATCGA	17700
CGTAACAGAA GACGACCGCT TAGATACAGG CAACCCAGCG GACATTGTTT ACACAAAAGA	17760
TCTGTTTTCT TTAGAAGAAA GTCCGCGCCT TGGTTGCGGT TTAATGGTGA TGAAAGATAC	17820
CACCTTCGAC TGGACACTAG AATACGATGA AATTGATTAT ATTATTTTCAG GACAATTAGA	17880
TATACTGATT GACGGACGTA AAATTTCTGC AACAGCCGGC GAAATTATTT TAATTCCGAA	17940
AAGTAGCAAA ATTCAATTTT CAGTGACGGG TGATGCCCGC TTTGTCTACG TGACTTATCC	18000
AGCAGATTGG CAAAGCCAAT AATTATTTTA TCGGCAGTCA GAGAATAATT TTGCCAAAAC.	18060
CCTTTACTCA GCGTTATAAT AAATAAGAAC GAAAAAGGG GTTGAGTGAA TGACATTGA	18120
GAATCCAAGT CAAAATCAAA TTTTGTATAT TTTAAACAA GCAAAAATA TTGCGTGGT	18180
TGGTTTGAGT AACAAAGCCG ACCGCACAAG TTACAAAATT GCGGAAATTT TACAGCAACA	18240
TGGCTATCGT GTATTGCCTG TCAATCCAGT TTTAAAGGT GAAACGGTGC TTGGCGAAAC	18300
GGTTTATGGG ACATTAACAG AGATTCCTGA AAAGATTGAT ATTGTTGACG TCTTTCGGCG	18360
GAGTGAATTT TTGCCAGAAG TGGCCGCAGA AACATTGAAA ACAGAAGCGC CTGTCTTTTG	18420
GGCCCAACTA GGTTTGAAA ACGAAGAGGC AGCGAACTTG TTATTAAATG AAGTCGTCA	18480
AGTTATTATG GACCGTTGTA TTAATTAATGA ATTGGCAAAA ATGGCAGATG CTGAATAAAC	18540
ACGTGTAATT TTATAAGAA AGAGCGTCCT GAACGGAAAG CTCTTTCTTT TTTGACCCAG	18600
ATAGAGTATT ATTGTAATGA TGGTTTGTAGA AGGAGCGAAC GGAAACGCA AGAAACCGGC	18660
TCACTCTTCT TCTAATCTT TAAGrraTAG TTAATGTACG CTTATTAAT AAACGAGAAT	18720
CTArGTATAA GGAGTTTTTT CTTTGGCTAA AAAAATTAAC AATGAGTACA ATGATGCCTC	18780
TATCCAAGTC CTAGAAGGAT TGGAAGCAGT GAGGAAAAGA CCTGGTATGT ATATCGGTTC	18840
AACCGATAGC CGTGGACTCC ATCACTTAGT CTATGAAATT GTCGATAACG CAGTCGATGA	18900

AGCGTTATCT GGCTACGGGA ATGAAATCAA CGTAACTATC CAAAAAGATA ATAGTATTTG	18960
TGTTGCAGAC TCTGGCCGCG GAATGCCAAC AGGGATGCAT GCTTCTGGTA TCCCAACAGT	19020
AGAAGTAATT TTTACTGTCT TACATGCGGG TGGTAAGTTT GGCCAAGGTG GCTATAAAAC	19080
TTCAGGCGGA CTCCATGGGG TAGGTGCCAG TGTTGTTAAT GCCTTATCTA AGTGGTTAGA	19140
AGTTCATATT GTTCGTGATG GCGTAGAATA CATGGAACGA TTTGAGGACG GCGGTAAACC	19200
AGTTGGTACC TTAAAAAAA TCGGCAAAAC AAAAAACGC AACGGTACCT CCGTTACTTT	19260
TTTACCTGAC GACACCATT TCTCTACGAC AAATTTTTCT TACGAAATTT TAGCAGAACG	19320
TTTGAGAGAA TCAGCTTTCT TATTAAGG CGTCAAAATT ACTTTAACAG ATGAACGTGG	19380
GGAAGAACCC AAAGAAGAAG TCTTCCACTA CGAAGAAGGT ATTAAAGAAT TTGTTGCCTA	19440
TCTAAATGAA GAAAAAGATA CATTAAACACC CGTTGTTTAT TTTTCAGGAG CCAAAGAAGG	19500
TATTGAAGTT GAATTAGCTT ACCAATACAA CGATGGCTAT TCTGAAAATG TTTTATCTTT	19560
CGTTAACAAT GTCCGCACAA AAGATGGCGG TACCCACGAA gTCGGTATGA AAACGTCAAT	19620
GACGAAGGCG TACAATGAAT ATGCACGGAA AGTGGGACTA TTAAAAGAAA AAGATAAAAA	19680
TTTAGAAGGT AGCGACTTCC GTGAAGGCTT AGCAGCCGTC CTTTCTATTC GAGTACCAGA	19740
AAACTTACTA CAGTTTGAGG GCCAAACCAA AGGAAAATTA GGTACACCTT TAGCAAGAAC	19800
CGTTGTCTGAT AACGTCTGAG GCGAACAAAT GGGCTTCTAT TTACAAGAAA ACAGTGAAAT	19860
GAGTCAATCA TTAATTCGGA AAGCGATTAA AGCCCGTGAA GCCCGTGAAG CGGCTCGTAA	19920
AGCCCGTGAA GAAAGCCGGA ACGtAAAAA CGGAAAAAAG GTGAATCGTT ACTTTCTGGA	19980
AAATTAACAC CGGCTCAATC CAGAAATCCT AAGAAAAATG AACTATATCT TGTCGAAGGA	20040
GACTCGGCCG GTGGCTCAGC AAAACAAGGG CGCGACCGGA AATTCCAAGC CATTTTACCG	20100
CTACGAGGAA AAGTCATTAA TACGGAAAAA GCGAAATGC AaGACATCTT GAAAAACGAa	20160
GAAATCAATA CTATGATCTA TACGATTGGC GCAGGCGTTG GTCCTGAATT TTCAATTGAA	20220
GATTGTAATT ATGATAAAAT CATCATCATG ACCGATGCGG ATACCGATGG TGCCCATATT	20280
CAAGTGTTC TATTAACGTT CTTCTATCGT TATATGAAAC CATTAATTGA AGCCGGCAAA	20340
GTCTACATTG CCTTACCACC ACTTTATAAA GTCTCTAAAG GAACTGGAAA AAAATCAGTC	20400
ATTGAGTATG CTTGGACCGA TGGTGAATTG GCAGAAGTGA TTGATAAAGT TGGTAAAGGC	20460
TATATGCTCC AACGCTATAA AGGTCTGGGG GAAATGAATG CAGAACAAC TTTGGGAAACA	20520
ACGATGGATC CTGAAACTAG AACCTTAATT CGTGTTCCGA TTGATGATGC AGCTCAAGCA	20580
GAACGCCGTG TGACCACATT AATGGGCGaC AAAGTCgAAC CTCGTCGAAA ATGGATTGAA	20640
CAACATGTTT AATTTACGTT GGAAGAAGAT GGCAGTATTT TAGATCGTTC AGAAGAAGAC	20700
ACTTCTGCAC CAACAGGAGA AAGTCTTTTA GATGCGGAAA AAACAAAAGA AGCAGAACAA	20760
ACCGATGACA CTGAAATCAG TTTGTTTGAC ATAGAGTAGG GAGTGACAAT TTTGGAAAAA	20820
CGCCAAGAAG TTCAAGAATT AACCCTTGAA GAAGTAATGG GCGATCGTTT TGGAAGATAT	20880

TCCAAATACA TCATTCAAGA ACGGGCCCTA CCAGATATTC GAGATGGGCT AAAACCAGTA	20940
CAACGCCGCA TTCTCTTTTC AATGAATAAA GACGGCAATA CCTTTGATAA AGGCTTTTCGT	21000
AAATCAGCAA AATCTGTCGG AAACATTATG GGAATTATC ATCCCCATGG CGACAGCAGT	21060
ATTTATGAAG CGATGGTGCG TCTAAGTCAA GACTGGAAAT TACGGGAAGT ACTAATTGAA	21120
ATGCAC	21126

## (2) INFORMATION FOR SEQ ID NO: 24:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3222 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

nAGAGAAAnC CGGTAGCTTT TTACGGTTTT CTCTTTTTTG ACTTCAAAG TCAAATCAGG	60
TACGACTTCT TTTTAAACAG ATAAGaATGT GGTATCATTG TAAAGGACTA TATAATAAAG	120
GAGATGGACG TTTTTTGACA GATAATCAGT CAAAATCTTT GGAATTAATA CTTACTGATT	180
CTGATGATGC ACACATGCTT TTAGGAACAC ATGATAAACA TATCAAATTT TTAGAAGAGA	240
ATACTCACGT AACGATTAAT AGTCGTGGCG AAGTTATTCA GTTAATTGGC GAATCATCAG	300
AGGTAGAGTT AGTCGCTTCC GTGTTAAGAG CGCTACAAAC ATTAATTCAA CGAGGCATCA	360
AAGTGCATAC GCCAGATGTC GTTCTGCTT TAAAAATGGC CAAAGCTGGA AATTTAGAGG	420
CGTTTATTGC CATGTATGAA GAAGAAATTA TGAAGGACCA TCATGGTCGG GCCATCCGAA	480
TTAAGAATGT TGGTCAAAAA AAATATATCG ATGCAGTTAA AACTCATGAC GTCATTTTCG	540
GGGTGGGACC TGCGGGGACA GGAAAAACCT TTTTAGCCGT GGTGATGGCA GTAGCAGCTT	600
TAAAAAAGG CGAAGTACAA AAAATTATTT TAACTCGTCC AGCGGTGGAA GCTGGCGAAA	660
GCTTGGGCTT TTTACCAGGT GACTTAAAAG AAAAAGTCGA TCCTTATTTA CGTCCAGTTT	720
ACGATGCCTT GTATCAAATT TTTGGTATGG ATCATACAAA TCGTTTAATG GAACGCGGCG	780
TTATTGAAAT TGCCCCGCTT GCTTATATGC GTGGTCGTAC ATTGGATGAT GCTTTTGTCA	840
TTCTAGATGA AGCACAAAAT ACCACGGTTG CCCAAATGAA AATGTTTTTA ACTCGTTTAG	900
GGTTTAGTTC AAAAATGATT GTTAATGGAG ATACTAGTCA GATTGACTTG CCACGTGGTG	960
TAACAAGTGG TTTAGTTAAC GCCGAGCGTA CGTTAAAGA CATTGAAAAA ATTGCCTTTG	1020
TAAATTTTGA AGCCAGTGAC GTTGTGCGTC ATCCTGTTGT AGCACAAATT ATTCAAGCCT	1080
ATGAAAAAGA ACAACAAAAG CATTCTAAG CAGTTAGGGA AGGAGTGAAA GTACCTTTTG	1140
AAAAAAGTAC TTACAAAGTT AGGGGATCGT TTAGGTAAAG CCTATTTACC TATTTTATTA	1200
GCCGTTTTTT CTCTTCTTTT ATTTATGATT ATGTTTGCA GCGTCCATCA GAAAAGAGTA	1260
GAAATTAAAG AAGGTCAACT GGCCGAGAAA ACCAttCGTG CAAATAAAAA TATCGAGAAT	1320
ACCTACGAAA CAGAACAAAG AAAAAAATTG GCAGCAGAAG CTGTCACACC CGAATATATT	1380

TATCAAGAAG ATACGGCATC TGTCCAACAT AATCGTATTG ATAAATTATT TAAATTGATT	1440
GATTCTGCGA ATGAAAAAGT AGACAAAGAG TACAGCAACA AGCAAGCGAA GGCCAAAAAG	1500
GAAGAAACGA TTCCTGCGCC GACCGTGGAA GAACGAGTAG CTAGCTTAAA ATCaTTGTTT	1560
GAATCTTTAC CGCAAGATGA AGTTACGTTT TATCaAAGTT TTCCTAATGT TTTTATCaA	1620
ACAATCTTTA CATTAACCTC TGAGCAATTG GATAAGGTTC GTTCAGAAAG TTTGATGTTG	1680
GTGGATGACG CGATGCAAAA TCACGTTCGA GAATCTGATT TGGATAAAAT TCGCCAAGAA	1740
GCTAACGGTA AAATTCAATA CTTAGATATT ACGAGTACCA TGCAACAAGT CATTGCTTAC	1800
ATTGTTAACC AAGGGATTGT GGTAAACGAT ATTGCCAATG AAAACGAAC AGAAGAgctA	1860
CGCCAAAAAG CGATGAATGC TGTTCAACCC GCAATGATTT ATCAAGGCGA AATTATTGTT	1920
CGTGAAGGTA CTCAAATTGA TGCAAAAGCG GTCGAAAAAC TTGAGCTACT AGGAATGACT	1980
AGTCAAAATA CCTCTATTTT TCCGATGGTG GCATTAGCAT TAGCTATTTT ACTTCAAGTT	2040
GAAGTGTTAA TTTTCTTTAC AAAACAAGTG ACTGAGCCTT CGCGACAACG CTCGTTTATT	2100
ATCTTTTACA CGGGTGCTAT GCTAATCAGT GTTATTTTAA TGAAATTCTT CCAGATTTTC	2160
CAAACAGAGC AGTTGATGTA TATTCCCTTG TTCTATCCGG CAGCCTTTCG GCCGCTGATT	2220
CTAAATCATT TTGTAAATCG TCGTTCAGGG ATAATCGCTG CTATTTTCCA AGTCGTTTTT	2280
GCGTTATTTA TCTTTTATAA TTCTATCGGC ACAAACCTCGT TGACGGTTAT TCTGATTATG	2340
TACTTATTCT CAGGATTTC T AGCAACAGTT GTTAAACGGA AACGGATGAG TGAGCAAGTT	2400
TTCCCAGCTT TAATGTGGGT AGTGGTCTTT CCTGTTTTCA TGGCGGTTGT CTTAATGATT	2460
TATCAAGGGA TGAGTTTAAAC AGATGGTAAA ACGTGGACAG CTTTAATTTG TGCAAGTGCA	2520
GGAACGGTAC TTTCATTTTT AGCAACAATG GGCTTGCATC CATATATCGA ATTATTAGTG	2580
ACCGATGATA GTATGATTGT CTTGAATGAA TTAAGTAATC CGAACCATCC GTTGTAAAA	2640
CAACTGTTAG AAGAAGCGCC AGGTACCTAT CATCATAGCA TGATGGTGGC TAGTTTGAgC	2700
GCTAACGCTG TTGCTGAGAT TGGCGGGCGG TCGCTATTAA CACGAGTTGC TTGTTATTAC	2760
CATGACATTG GTAAAATTAA ACACGCCAAT TTCTTTGTGG AAAACTTACC TGCTGGTGCA	2820
GAAAATCCAC ATAACTTTTT ATTACCAGAA GATAGTAAAC AAATTATTTT TGGCCACGTA	2880
ATTGATGGCG CCAAAATTTT AGAAGAGTAC AATATGCCGC AAATGGTGAT TGACATTTGT	2940
CGTCAACATC ATGGCACAAAC ATTGATGAAA TTTTCTATG TGAAAGCAAA AGAACGCAAT	3000
CCTGaAATTA AAGAATCCGA CTTCCGTTAT CCTGGaCCAC GTCCACAAAC aCGAGaAGCa	3060
GGGATTGTGa GTATTGcTGA TACTTGTGAG GCTGCTGTGC GGGCGATGGA TCATCCCACC	3120
AACGAAAAAA TTnCAAGCCT TTGTGCATAA TGTGATTCAA GACAGnATTT CAGATGGCCA	3180
ATTAGATGAA TGTGGGTTAA CGATGAAAGA GATTCGCATC AT	3222

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7812 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GGACATGATT ACGTTAGACT TAAACTTACC TAGTATGGAT GGGATGGCTG TTTGTCGTGA	60
AATTCGCAAA GTTTCAGCCA ATGTTCCCAT TATTATGTTG ACCGCTCGTG aTTCTGAAAG	120
TGACCAAGTA ATTGGCCTTG AAATGGGTGC GGATGATTAT GTGACAAAAC CATTTAGTCC	180
ATTAACCTTG ATTGCACGGA TGAAAGCTCT ACATCGTCGT GCCGAGGTGG CAGAGGCTGC	240
GCATGATACG TCTGAGAATA CAGATGAAAC ATTTGACGTG ATTACGGATC ATTTTAAAAT	300
GAATACGAAA ACACGAGAAA CATATTTAGA TAATCAATTA ATTGAAGGCT TAACACCTAA	360
AGAATTTGAT TTGCTATACA CTTTAGcAAA AAAACCACGC CAAGTCTTTT CACGcGAACA	420
ACTACTAGAA TTGGTTTGGG ACTATCAATA CTTTGGCGAT GAACGGACAG TTGATGCCCA	480
TATCAAAAAA TTACGACAAA AAATTGAAAA AGTGGGTCCT CAAGTCATCC AAACGGTTTG	540
GGGCGTAGGG TATAAATTCG ACGATTCAGG TGTTCGTTAA TGAAGTATTT GTATCAACAA	600
TTACTTGCTT TTATAGGTGT GATCGCTCTG ATTATTCTCA TTGTCGGAAC GTCTTTCACA	660
CAATTGACCA AGCGCACAAAT GCAAGAAAAT AACTATGAAC AGCTGTATGG TTATGCAGAA	720
TCGGCTTTAG AAACACGTGA CTTTTTTATT AATGTCGCAG GCGTTTCAGA TCGAGATGTT	780
TTATCCTATT CATTtCAGTT GACTGAGCGC GTACTCCAGA AGCAAGATGT GCAATTTGTG	840
TTTATTAACA AGGACCGCGA AGTGCAGTAC CCACCAAGTTG ATAGCACAAA AAAATTAGAC	900
TTTTCTTAA TCGATAfAA TTGGGATCAA ATCATGAAAG GCAATCGTGT TaTGCAACAG	960
AAAACATTGA TATTTACGGG GCGCGAAATA CTTCGTCTTA TGTAATGTTA CCGGTGTATG	1020
CATCAAATCA GTCTTCTGAT AAAAAAGTGA TTATTGGTTC ACTAGTTATT ACACAACCTG	1080
CCAAAAACGT TGACCGAAGT GTTCAGTCCG TGACACAAAA CTGATTAAA GGCTTTATTT	1140
TTTCTGGTGT GATTGCCTTG TTAATAAGCT ATCTATTTGC GACTTCCAA GTGAAACGAA	1200
TAAATCGGAT GCGTAAAGCC ACCAAAGAAA TTAGTAGTGG AAATTTTGAT ATTCAATTAC	1260
CGGTTTCATGA CAAAGATGAA TTTGATGACC TAGCAGAAGA TTTTAATAAA ATGGCCGCGT	1320
CTTTAAAGA ATCACAAGAA GAAATCAATC GGCAAGAAGA GCGCCGTCGT CAATTTATGG	1380
CGGATGCCTC TCATGAAATG CGGACGCCTT TAACAACTAT TAATGGCTTG TTAGAGGGAT	1440
TGCAGTATAA TGCTATTCCT GAAAATCAAA AAGAGAATGC GATTAACTG ATGCAAAATG	1500
AGACCGCTCG CTTGATTGCG TTAGTCAATG AAAATCTTGA TTATGAAAAA ATTCGGACCA	1560
ACCAGATTCA AATTGTTGTG AAGAAGTTCA ACGGAACAGA AGCTTTAGAA AATATTGTGA	1620
CTCAACTAAC GGCCAAAGCT GAAGCCGCCG GTAATCAATT GTATTTAGAC ACCACCGAAC	1680
CAATTGATGT GTATGCCGAT TATGATCGTT TCGTTCAGT GGTGTCAAT ATTGTTCAAA	1740

ATGCGATCCA ATTTACTGAA AATGGCGAAA TTCATATTGC TTTAGAAAAA GGCTATTTAG	1800
AAACCATTGT CCGTATTTCT GATACAGGAA TCGGCATGAC GGAAGAGCAA ATTCTGAATA	1860
TCTGGGATCG TTATTACAAA GTAGATCCAT CCCGTAAAAA TACGAAATAT GGAGAATCAG	1920
GATTAGGCTT ACCAATCGTT CAACAATTGG TCCGTTTGCA TAAAGGAAAA ATTAATGTTG	1980
AGAGTGAACT TGGTAAAGGC ACCACGTTTA TTATTTCAAT TCCTGATGTT GAAATTACTG	2040
AAAATAAAG AAAACGTCTG GTCACAATCA GTGCCAGACG TTTTCTTTAG TTTAATAACG	2100
CATGTAGAGA CAGGGGGTGC ACATAGGGAC TATCAATAAT TTGTGGATCT TTTGTCAGAA	2160
ATAGCAAGCG CTTGTTTCTT TTTTGAAGTCT CTATTTTTAA TTGAGGTAAA ATTTCTTGAC	2220
GTTGTCCAAT CGTTAATGAA GAAAAGATGT CATCAATCAC TAACGTACTT GTTTTTGAAA	2280
GCAAAAGTTG TATTATGCGC ATTTTAATTT TTTCAAAAGT CGTAAAAGAA GTAAAGGAAC	2340
GACCAAGTAA GGCAGGTTCT AATTGGAACA CCTGGAAAGC TTCGGCTAGA ACTACTTTTC	2400
GATCCCGTTC TTTGATCGTA CTAGCGATAA ATAAATTTTC TTCTAATGGC AAGTAGGGCA	2460
AAAAGGTTTC ACTAGAAGAA AGCCTAGTAA TGGTTAATTG GTTTGTTTGG AAAAAATCGG	2520
CCAATGCGAG TTCTTGTCTT GGAGCAGAAT AAATGATTCC ATAACTTCT CGCAAAGctA	2580
ACGCGTCGAA TGAGGATTTT TTTTCGGTTA TTTTTTTGAA TGATGGTGTG ACATTCATAG	2640
AAAATCCCTC CCTTATTTCT TATGTGTACG TAAATAAATC AACGTGAAAG GAGTCCCGCA	2700
AACAAGAGGA ATGCCAAGCA AGATTAAGCA GTTAACAAAG CAATTACGGA TTGTATAAAG	2760
TGACTGCATA CTTTTAAAAT CAAAAGAAAA CGCTTGGATC GTTTGAATCC CTAAATGCTG	2820
GATATTTGTA ATCGTGACAT CCTGGTCGGT ACTTTTAGCT AAAAATGCTG GAGTTTCCAA	2880
ATGAGACCAC TGTAATAATAG CATCACGAGC AGCCCCAAGA ACAGTATTAT ACGTTTCTTG	2940
AAACATTAAA GTGAAAAGTA AAAAAGTAA AATTCCCTAAT AGAGCCGGTA AACATAACTC	3000
TAGTAAGAGC TGTTTGGTGA CATAATATT TGAAAAGTGC ATGTGATACC ATGTATGTAT	3060
TTCTTCCTTT TTTATGCGTA ATGTCACATA CATCAACAGA AGAAAAGTAA CACCAAAGAA	3120
AATGACACTA CCTAGATAGA GGTGCCAATA GTATTGGATC AGTTGGTAAT GGTAAGTAGT	3180
TTGTTTTTCA ATATCGACAA AGTCTGGCAT CGTTCGTAAA AACTCTTTTT CAAAATCAAT	3240
TAAGTTAAAC AGCCCAGTCA ATAAGAAAGA GAACAGGAAG AAGAAGAGTG CGGTGATTAA	3300
CGTGATTTTT TTATGATAAA AACTACTAGC TGTAGCGTAA CGGACACTTC TCATCAACTT	3360
TCACCTCTCT AACTATCTT TAGTATAGTG aAATAATATG AATTawTTak GaaAAAAGtAG	3420
CAAAAaGCTT GAGCAAATkG CTCAAGCTTT TTGAAGATTt ATTTACCCGT TTTtGATAG	3480
CCTtGATAAC TGTCAGtGTG ATATtCATCA ACGGtGATT TATtATkGTT TTTACTGTAA	3540
ACACTCGTTG ATTTTnCAC nTTTcTTTTT CAATTGCTTC CAGTTGTTGC AATTGGTTCT	3600
TGTAATCATA GTCTTCTGGG TTAAGTGGT TTAAGCCGCT ATTGGTATAG AAACGAAGCA	3660
AGTCACCGTT AGTAATTTGG TCAGACGTTT CTAATTGTTT CGTCGCTTTG GCTTTTAAAT	3720

CAGCGACTTC TTTCTTAACT TCTTCGGTTG GCTCTGTAAT TAAAGTGCCT GTTTTAGTAT	3780
CGTAAATGCT AGAACCTAAA ATAGTATATT TAGGTGTTAC AACATTGCCG TTGCGGAAAG	3840
CAACGATTTG ATTGTGTTGT TTAGAGAATA AATCTTGGCC TAACTGAATG TAGTTTTTCG	3900
TATCGACACC AAGCAAGTGT AGAAGGGTTG GTAAGGCATC AACTTGGCCG CCATACGTAT	3960
GGTTCACACC GCCATTTTCT TGACCAGGAA CGTGAATCAT ATAaGGAACA CGTTGCATAT	4020
TGGCGTTGTC AAAATCGTTC CAATCTGCTT TGGTTTTACC TACTAACTCT GCTAAGTTTT	4080
GGTTTCTTGA ATTAGAAACC CCATAGTGAT CGCCATAAAG AACAATGACT GAGTTTTCAT	4140
AGAGACCTGA AGATTTCAAG TAGTTAAAGA ATTCTTCGAC CGCTTTATCT AAATAATTCTG	4200
CTGTTGCAAA GTAACCATTA ATTGTTTCGT CAGAGGTTTT TGCAATTGGG AATCCTGCTT	4260
CATCGTTTGT AAATTGAGAA TAAGGATAAT GATTGGACAC AGCAATAAAT TTAGAATAGA	4320
ACGGTTGTTG TAAATGTTCT AAATATTGAA CGGATTGATT AAAGAATGGT TTATCGTGTA	4380
ACCCATATTG GAATGAGTTA TCTGAATTAA CATCATAATA ACTAGCATCA AAGAAATAGT	4440
CATAGCCTAA ACGTTTATAG GTTTCATTTC GATTCCAGAA GTTACCGGCA TTTCCGTGGA	4500
AAGCAGCGCT AGTGTAACCT TGC GTTTGTT TTAATAATC GGGTGCTGCT TCAAACGTAT	4560
TTTTGCCACC AACTTGTGTA AAGAGTGAAC CTTGGTCTAA ACCAAATAGT GAATTTTCAA	4620
GTAACGTTTC GGCATCACTT GTTTTTTCCTT GACCAACTTG GTGGAAGAAG TTATCAAAGC	4680
TAAAGTACT TTTACTATTG TAAAGACTAT TGATAAATGG TGTGACTTCG TGTTCTACGC	4740
CATTTTCATC TTTTAATTG TAATTAACCT AAAATTGTTG GAAACTTTCT AAATGGATAT	4800
AAATTACATT TTTGCCTTTT GCAATACCAA ACTTACTATC ATCGGGTGCC GCATATTGTT	4860
GTTCACGTA ATCTTCAACT TGTTTCATAT CGTTAGCACT GGCTTCGGCA CGGACTTGGT	4920
TTGTTTGTA AGTCGTAATG CCATCGTACA CGGTAAAGGC ATTTAAGCCT AAGAATTTCA	4980
CGATATAGTC ACGAGAGAAT TGACGTCCTA ATAATTGGGG ACGTTCGTGTT TCAGCCATAA	5040
ATAGATTGAA TAAGAAAAAG ACAACAGAAA GGAGTGTGAC AGAGACTGCT ACACGTGCGC	5100
GGACTGGACG CGCATCTGTT TTAATTTTTT TCGTTAATAA TAATACGCCA ATAATAATGA	5160
AGTCTAAGAA ATAAATGACG TCATAAGGTC TAAAGAGACG TAAGGCACTT TCGCCAAGAC	5220
CGCTAGCTAC TTTGCCAGCA CCAAGCATCG TATTAAGTGT GATGAAGTCA GTAAATTCTC	5280
GATAATACAC GACATTTGAG AAGAGCAGTA AACTCATCAA GAAGTAAATA ATCATCATTG	5340
TAATATAAGA AGCTTTTTTA CGGCGAACAT ACAAGCGAT GGCTAATAAA AAGACCGTTG	5400
TGGCAATTGG ATTAATTAAG AGTATGAAGT ACTCGTTAAT ATTTTCCAAA CGTAAATGGA	5460
AATCAACGGC GTACGCAAAG AGGTTTTTTA GCCACAGTAA GATGGCGATT AGCGAGAATA	5520
ACCCAGTCG CGTGTTCAAA ATATTTGTTT TAAATAATTT TTTCAAATA AACGTCCTTT	5580
CTAAACGATG ATTCTATAAA AGTAAATTA GGTAAATTC GTATTTTACA ATATCGCTTT	5640
AATTTTACTC TTTACAAAT TTTATGTCAA TTCTTACCGA AGTGAGGAAT GGTGATGTTA	5700



AAGAAAAC TT ACTATTTCTT TAAACTTTTT AGTCGTAA ATGACGTGAA GCTTGAATCC	5760
TTTAAAGACA GATGGGCCTT TTTTGCTAT ACTAGATAGG ACTTTGAAAA GTAGGTGTAC	5820
AGAATGAAAC TTCAAGTAAC TAAAAAAGCA GAACATAAAT TAAAAAAGG ATATCCTTTA	5880
ATTCAGAAAG AAGATTTGCA ACAAGTGCCA GCGCCGTTGC CAACAGATTG GCTTACTTTG	5940
ATTGATAGTA AAGGCCAGCG TTTAGCCGAA GGATATTTAG GCGAACAAAA TAAAGGAATT	6000
GGTTGGTTGC TGTCATGGCA TGGACCAATT AATCAATCTT TTTTCAACA ACTCTTTGAA	6060
ATAAGTCGTG AGAAGCGGAC GTCTTTTGAA AAGGATTTCG TAACAACCGC GTATCGATTA	6120
TTAATGGAG AAGGGGATGG CATTGGGGGG CTAATCATCG ATCGCTATGC AGATTACGCC	6180
GTTTTTTCAT GGTATAATGA AACCTTTTAT CAGAAAAAAG CCGAACTTTT GACGGCGTTT	6240
CGCACAGTTT ACCCTGACAT TATTGGTGCC TACGAAAAGA TTCGCTTTTC CACAAAAGAT	6300
TTACCAGAAT CACAGTTTTT GTATGGCGAA CAAGCACCCG AGCCATTACT TGTCACTGAA	6360
AATGGCGTTC AATTGCTAC CTATTTAAAC GAAGGATTAA TGACAGGCAT CTTTTTAGAT	6420
CAAAAAGAAG TTCGTGGTCG ATTAGTAGAT GGTTTTGCAG TAGGTAAAAC AGTCTTAAAT	6480
ATGTTTAGTT ATACAGGTGC TTTTTCAGTA GCTGCCGCTA TGGGGGGTGC CGTGGCCACC	6540
ACTAGTGTGG ATTTAGCTAA ACGAAGCTTA CCGAAAACGA CTGAGCAATT TGAAGTCAAT	6600
CATTTAAATC TTGCCCGCA AAAGrTTATT GTTATGGATG TTTTgACTAC TTAAATATc	6660
TTCCGCTAAA GGCTTGmGTT ATGACATGAT CATTTTGGAT CCGCCGAnTT TTGCTCGCAA	6720
CAAAAAGAAA GTTTTTTCCG TTGCTAAAAA TTATGGGGAG CTAGTGAAAG ATTCTATTGA	6780
TATTTTAAACG GATAAGGGAA CGCTTATTGC CTCAACCAaT GCGGCAAATT TATCCTTAGC	6840
CAAATATCAA AAAATGGTGA TTAAGTGCCTT GCAAGAGAAA AATGTCCGTT ACAAGATTAC	6900
GGACACGTAT CAATTACCGG CAGACTTTCA AGTGAATCCT AATTTTCCAG AAGGTAATTA	6960
CTTAAAAGTC CTGTTTATCG AAATTGAAAA ATAATTTTTT ACACCAAGCA AAAGTCCCAG	7020
CGCCGCTAAT CGGTATTGGG ACTTTTGCTT GGInTTTATT TTATAGTAAC TTTTGGGCTT	7080
TTTGAGTGTT TGCAAAATGc AGTTCGCGTA GTTTGCTAAT AAGTCTAGGC CGCCTTTTTT	7140
CAAAATACGG GCGGCTACTT GGTCTGATTT GCTGCCAGCA CCAGAGGGCA CAGAGAAACC	7200
CAATTCTGCA GATTCTTTGT CTAATCTTC TAAAAAAGCA GCGCGGCTAA TAATAGAGGC	7260
GGCAGCAACG GCAACATGAT ACTGCTCTCC TTTAGTGACG AAGAACAATT TTTCACTGAC	7320
TTGATTTTTT TCATTACGAA CATATTTCCG ATAATTATTT TCAGGGGTGA ACTGATCAAT	7380
TAAATCCCT TCTGGCTTTG TGGGAGCCAA TTCTGTAGC AACAGATAAA TCGCTTGTT	7440
ATGCAAAGCG ACTTTCATGT GTACAGCATT GTAGTTTGGT TGGATCTCAT TGTATTTTTT	7500
GGGTTCAACG ATTAATAACT TGTAAGGAAT CAGTTCTTTA ATGACATGTG ACAATTGAAT	7560
AATTTGTGGG TCGGTTAATT CTTTGGAATC ACGAACACCT AATGATTTTA GTTTGCTAAT	7620
CATGnTTTTG TCAACATAGG CAGCACAGAC AGTTACTGGA CCAAAATaGC TACCATTaCC	7680

GACTTCATCG GAACCTATCA CAGgACcAT TTGCCCCAAA TCCAGCAGGG CAGTTGGTAG	7740
TTAGTCTGGG GGAAGAACT TTTTnTCTTC TTGGGTGTG GTTACCTTCG GTTCCCTTT	7800
CCCCAAAAC CG	7812

## (2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1769 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

TTCAGCGACA GGGATGTTAT TGATGATTAA ACGTTATCAA TCGCAATTCG TGTTTGGCT	60
ACTAGGAAAT ATTTTTTCAA TTTTGTTATG GTTCCGAGCA GGGACGCACG CTGGTGGTGA	120
TTATGCGATA TTTGTTATGT ATTGTATGTA TACTTTTAAC TCAATTTTGT GTATGTTAA	180
TTGGTTGAAG ATAAAAAATA AAATGGAAAA AAATAGGTAG GGAGTTCGAT GGCAGTgATT	240
GTTTTAGCAG GTACGATTGG CGCTGGTAAA TCTAGTTTAA CAGCTTTAAT TGCAAACCGT	300
TTAGGTTTCA GAGCCTTTTA CGAATCAGTT GATGATAATG AGGTATTACC TTTATTTTAT	360
GCGGAACCGG AAAAGTACGC ATTTTATTAA CAAATTTATT TTTTAAATAA ACGTTTTGAT	420
AGTATTAAAC AAGCATTAAc GCATGAAAAT AATGTATTGG ATCGCTCGAT TTATGAAGAT	480
TCATTACTTT TTCATTTAAA TGCAGATTTA GGTCGTGCCA ACGAAACAGA GGTCAAAGTG	540
TACGATGACT TATTACAAAA CATGCTCCAA GAATTGCCTT ATGCTGCGCA TAAAAAAGA	600
CCTGATTTGT TAGTTCACAT TCGTATTTCA TTTCCCCAAA TGTTGGAACG CATCAAAAAA	660
CGTGCGCGTC CTTATGAACA AATTGAGACA GATCCAACGT TGTATGATTA CTATCAAATG	720
TTGAACGAGC GTTATGATCA ATGGTATGAA GACTATGATG AAAGTCCTAA AATTCAAATT	780
GATGGCGATA AATATGATTT TGTCGAAGAT CCTGAAGCGT GTCAATATGT CTTGGCGTTA	840
ATTGAGAAGA AAATTGAAGA ATTAGAAGA TAAATAAAAA GAAGCAAGTC CGGACTTGCT	900
TCTTTTGTAC TTAAATGTG GATAGATAGG CTAAAGTTGC TACTAAATAA AGGAGTGTAG	960
TGCTAAATTA TAGAATCGGA TTTTAACCTT CCCCTAAGTT AAAATCGGAT AATCTATCTA	1020
TCGACAGTTT AAGTTTGTAT AAAATAGCGG TAGCTATTAT CAACGAAGTC ATTCATCGGT	1080
TGACCTGACT TCTCTAGAAG TATACCATCG AAAATAACCG ACAGTTGTTC AAAATTGGA	1140
CTTAGTAAGT GTAAATTTTG ATTGAAAGTT TCTGTAAATC AATCATTTAT TCAAAAAATA	1200
ATGCTCGAAA GAAGGAAGAT AAATGAGCCA AATTTGGTTT GAACGCATTC TACGTTGGCG	1260
CAAAAAAGCG CAACGATTTT TACAAGGACG TTATGCCCGG TTCGATGAAT TGAATCGAAC	1320
GCTGTTGCTA ACAAGTGTCT TGcTGGCGCT CATCAATATT TTTACTGGCT ATCTTTGGGT	1380
TCGGTTATTG ATTTTAATCA CAGTCGCTA CGTGTATTAT CGCTTTTTTT CAAAACATAT	1440
TCACCCGCGA TTGAATGAAA ATCAATGGTT TATTCTTCGA AAACAACGCC TAATGCAAAA	1500

379

GGTTCATGCT TTTCCGACGC SACATAAGAC TCAGAAAGAG TATCGTTATT TTAAATGTCC 1560  
 AAAATGTGAG CAGTCATTAC GTCCGCGGAA AGGACCGGGA ACAATCAAAG TCACTTGTTT 1620  
 AAACTGTCAA AATCAATTA TCAAAAAGT CTGATCTTCT GTGTTTTTTG GTCTAGGCAA 1680  
 CTTTCGGACT TGTGGTATAC TTGAGAAAAG AAACCAATGG ATATTCTAAA AGGGGAATAT 1740  
 CTTGKTAAAA AAGGAGCTTA AAAACAATG 1760

## (2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1908 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CnGTGCTTGA CGCATCAACG CCTCnCTCGA GTATATGnGT TCACCTTCAC CGCGTTTAAT 60  
 AGcAAAcGcT GTTaccGTTA AATTTCATT GGCATTAGCT AATCGATTG TTACTTCTGC 120  
 AATCGTTGCT TCGACATTTT CTTTTCTTTT GGCAACATA AGACTGACAG AAGACTTATA 180  
 GGTAATCACC GCAAAATGG TAAAACTAC AATATAAAG AAAGAACTTG CGAAAGCCCA 240  
 CTTTATAGTT AAAGATGGG CTTCCAGTTC TTTTTTATC GTTCTTTTCA TTAATTAGTT 300  
 CTCACGAACG CATAACGTAG CCACTCCAC GGACAGTTG GATGTAGCTT TCTTCTCCAG 360  
 GTACGTCAAT TTTATTCTT AAGTAGCGGA TATACACATC CACAACGTTT GTTTCTACTT 420  
 CTGTTTCATA GCCCCAACT ITATTTAGTA ACACATCAG TGCCAAGACA ACGTTCACGT 480  
 TTTCCATTAG CGTTAATAGT AATcGTATC GCGkTTKgtt AATTCAATCA TTTCAGAATT 540  
 ACGACGAACG ACACGATTTT CTTTTTCAAT TGTTAAGTCA CGATATGTAA TCGTTGTTTG 600  
 TTTTGCAACG TTTTATCGC CCTCAATATC AATACGACGA AgTAACGCAC GTAAACGAGC 660  
 TAACAATTCT TCAATTGCAA AEGGTTTAAc AATATAATCA TCCGCTCCAT GGTCTAAGCC 720  
 AGAAACACGG TCAATTACTG AATCAGCGC AGTCATCATA ATAATTGGTG TATTTTTCAC 780  
 TTGGCGAACA CGGCGACATA CTTCTAATCC ATTAAATTCT GGTAACATCA AATCAAGAAG 840  
 GATAGCATCC CATTCGTTGT TAAAGCGGC TTCCAATCCT GTACGACCAT TGTAGTGATC 900  
 TTCTGTGTA TACCCCTCAT GTTTAATTC AAGCTCAACG AATCTCGCTA AGTTCTTTTC 960  
 ATCTTCAATA ATTAAATGT IGCTCAATTG ATTAATTGTC CTTTCTCTTT AAAAGCTATA 1020  
 CTCAAGCTCA ACCTTTTTAT TCCTTACTAC TTATTAGAGA ATTACTACTC AATCTATCTG 1080  
 GTTCTTTTTT AAGCCACAGT AATACCATGA GCATGGTGAG CCACACTCAT GGTCTCCAGC 1140  
 GTGTTTATTC TTCGTGTGAC CAGCTGTAAT GGTAGATTCC TTCTTTGTCT GTGCGTTCGT 1200  
 ACGTATGAGC ACCGAAGTAA TCACGTGTG CTTGAATTAA ATTCGCTGGT AAGCGATCTG 1260  
 AACGATAAGA ATCGTATTA GCAATTCAG ATGAGAATGT TGGTACAGGA ACACCTGCTT 1320

380

GAACAGCaAT CGCaACAAC TCACGAACGG CTTGTTGATA TTTTTCGTA ATTTCTACGA 1390  
 AGTATTCATC TAACAACAAG TTTTCAAGTG CTGGATTTTT TTCATATGCA TCGGTAATTT 1440  
 TTTGTAAGAA TTGTGCACGG ATAATACAAC CTGCTCGCCA GATTTTTGCA ATTTCAACCA 1500  
 ATGGTAAATC CCATCCGTAT TCTTCAGAAG CTGCACGGAG TTGTGCAAAA CCTTGCAGAT 1560  
 AACTCATTAA TTTACTAAAAG TATAAaTgTT CGCGAATTTT TTCAATCAAT TCTTTTTTAT 1620  
 CGCCTGCAAA GTTAAAGGCT GCAGGTTTTG ATAAAATGCC ACTTGCTTTT ACGCGTCTT 1680  
 CTTTGTAAAG CGAAATAAAA CGTGCAAATA CTGATTCTGT AATTAATGGT AATGGTACGC 1740  
 CAAGATCTAA CGCACTTTGG CTTGTCCATT TACCTGTTCC TTTGTTGCCT GCTGCATCAA 1800  
 GAATAACATC AACGATTGGT TGGTCGGTTC CTTATCATC TTTACGTGTC AAAATGTCAG 1860  
 CCGTGATCTC GATAAGGTAA CTGTCTAATT CGCtTCGTTT CATTCTTT 1900

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7515 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

TCTTCTATkG TTAaTGcaw CTTGTCCAAC TCCTTGyCTT CTTCTACATT ATAACtAGT 60  
 TAGGTGACTT TTTCACTAAT AACCAATTGA AACTTTTTCT TGAAACTTT TTGCATTTTT 120  
 CATAATTTTG CGTTATTAGC GGATACATTA ATGAAGTTAT GGTAACATA CATTCTGGAT 180  
 GGTTCATTCC AGCCAGCCAA TTAATAACTA TTTTtagAGC GTTCAACACG CTTATTCAAG 240  
 ATTTTTTTAT AAAGGTGGAC TTATTTTGAT GAAAAACAAA TCGAAGCTTT TATGGTtAGG 300  
 AAGTTTTTTC TTGCCGTTTT TACTTCTGTT GCTTGATGAG ATGACATTAC AACTTGCGCC 360  
 TTTTGGTGAC AACAACTTAT TAGTTAGTGA TTTAGGCACA CAATACATGC CTTTCTTGAG 420  
 TTTTTTAAAA CGCTCTTTTC ACGAGGGAAT AACTACATTC TACTCTTTTT CCAATGAGAT 480  
 TGGTGAATCG ATTGTTCTT TGGCAGCCTA TTAATTACTG AGCCCTTTTA ATGTGTGGC 540  
 TTTCTTTTTT CTTATGAAC AATTACCTAT AGCAATTTTA TGGATTATTA CTTTAAAACT 600  
 TGCCTTGATG GGCATACGA TGTTTAGCTA TTTAAATAC ACCTATCAAA AAGTTGATGG 660  
 CACCACGTTA CTTTTTTCAA CGTCCTATAG TTTCTGTGGC TTTGTACCG TCTATAGCCA 720  
 AAACTTTATG TGGTTGGATG CACTGATTCT ATCCCCGCTT ATCTTATTGG GACTCCAACG 780  
 TTTATGGGAT CAACGCAAAT GGGGCTTATA CAGTATCACC CTATTTTtag CGATTGTGAC 840  
 GAATTACTAT ATGGGTTATA TGATTGTCT CTTTGCTGTT TTGTACAGTA TCTATTGGTT 900  
 CTTTAAAAAG AATACCAAAG CCCATGCTAT TCGGCAATTT TTTAAACAAA GTCCGTTATT 960  
 CATCCTCGTT TCTTTTTTAA CAGGAACCGC AACTAGCTTT TTATTATTAC CTGCAGCAGA 1020  
 AGGGATGCTC TATACAAAAA AAGCAGACTT TGATGTCTCG ACGTTCTTTT TAACGCCTAA 1080

GTTCAACACT TCATTCTTCT CGCAATTAGG CTTAGGTTCT ATTAATTATG AGTTGCGTCT	1140
AGACCATTTA CCAACCGTTT TTGCTGGATT ATTTGTGACA CTCCTCTGTG TTGCCTATTT	1200
TCAAACGAAA CAAATTGCGT TAAAAGAAAAG AATAGCTTCA GCAATTCTTT TATTCACTCT	1260
TTTTTTAAGC TTTTGGTTAG AAGCCTTTAA TACTGTCTGG CACATGTTTC AAAGCCCAGC	1320
TGGTTTCCCT TACCGAAATG TCTTTATTTT TAGCTTTTTA TTGATTGTCT TTGCCTATGA	1380
AGTTTGGCTT AAAAAAGTGA CCATTCCTTG GACCGCACCT ATTATTTTTA GTCTGCTGTT	1440
AGTCATTGGC TACGGATCTT TGTATTATGG TCCACAAAAG AATCTCTTAA TTTCGATCAA	1500
TTATTTATGG CTAAGTTTAC TTTTTATTG GTTGATTTTC TTTTGCTTGC GTTTAGCGCA	1560
CAAAAAGGCC TTAAGAACT ATGTAGTCGT GGCTTTATTC TTAGTGGTCA GCACAGAATT	1620
AACAACGAAT TTCTGGATTT CTTTTAAACA CATGCCCTTT GGTAGCCAAG CAACATTTGC	1680
CCAAGATTAC CGAAAACACA GCCAATTAAT CGATGAAAAA ATGGCCTCAG CACCAGAACT	1740
TTATCGCATG AAACAAGTCA TTCCTTCCAA AGAAACAGGG TTCCGTGAAA TCAATAACGG	1800
CTACAATAAT CCTTTACTTT ATGGCTATGC TGGCGTTTCT AGCTATACTT CGACTTTAAC	1860
TGCCACTACC CAAGACACAT TAAGTGCCT AGGTTTATAT CGAAAAAATG ATCGCCGGAT	1920
TGCTTATGTG GATAATTCAC AACTAACTAA TTTACTTTTA AATGTGAAAT ATGACTTTTT	1980
ACCAATTGAA AAACCTACCA GTGAAAAATT GCTAAAAACa GTTGGTTCCA CGAAAATCAT	2040
GGaAAACGAT GAAGCGATCG GTATGGGCTT TTTGGCGCCA ACCGCATTAA CCAAGTTAAA	2100
ATTAGCAAAA AATAATCCTT TAGATGCCCA AGAAGaACTC CTGCAAACGC TCGTGCCGAC	2160
GGATAAACCT TACTTTAAAA CAGCTTCGTT GATAAATGAA CCTCATCATA CCAACGAAAC	2220
AATTGAAGCA ACCTTTAAAG TGAATAGCAC TGGTGACTTG CATCTTTATA TTCCGAATTT	2280
AAAATGGAAA AAAGTTACAC AATTGAAGGT AAACCAACAA GTTATCTCGA CGCCGATTTA	2340
TATTGCAACC AATCAACTGT TCAATTTAGG GCATTTTGAA AAAGGAACTA CCGTGACCCT	2400
GTCATAACT GCTGAACAAG TGGTTGATTT AACCAATTGG CAACTTCAAA CTTTAGACCA	2460
AACAGCCTTT AATCGTGCGG TTGACAAATT ACGCCAACAA GCTCTTCACG TGAATGCAAC	2520
TAAAAAAGGC CATTTAAATG GCGCACTGAA TGTACCTGGA AATGACACTC AGTTACTGTA	2580
TACGTCCATT CCTTATGACC AAGATTGGCA AGTCAAATCG TCGCTACAAA AAGAACCCTT	2640
AAAAACACAA CGCATTCTAG GTGGCTTTTT AGCTGTTGAA GTCCAGCTG GCAAACAGCA	2700
GTTAACCTTT GCCTATCATC CAAGAATGAT TTATCTCGGC ACTGCCGTCA GCGGAACGAT	2760
TTTACTGGGA ACGGCTGGTT ATCTTGGGTT TAAAAAATAC CGTCGAAAAC GTCAGGAGGC	2820
CACTCATGAC TAATCAGAAA AAATTTTAA ATGGACTCAT TCGGTGGTTA CCCCTAGTTG	2880
GACTGGTGCT TTTCTTCGGC TTAATTCTAT GGGGATATTC CAGAGGAATT TTTCATTCTG	2940
TTGCTTCCTT GCAAGCTTTT ATTAAACAGT TTGGTAACTA TGCGGTTCTG TGTTTCATTC	3000
TTTTACAAAT CGTCCAAGTG ATTATCCCTA TTCTACCTGG AGGAATCTCT TCTGTTGCAG	3060

GAATGTTGAT GTTTGGCAAT TTGCAAGGAC TTTTGTATAG CTATCTTGGA TTGATTakTG	3120
GTGAGTTTAT TGGCTTTTTA CTGGTTCGTT ACTATGGACG TAGCTTTGTA AAAATAATCC	3180
TTTCACCCAA TAAATACAAA AAATTTGAAG AGATTTTGGA CAAAAATGAA CACAATGTCA	3240
AAAAGCTCTT AATCTTTACT ATGTTAGTCC CCTTCGCTCC AGATGATATT GTTTGTTTAG	3300
TGGCGGGAAT TACAGATATT TCTTTAAAAG AATTCATGAA AATTGTTCTT TTATTGAAAT	3360
TCTGGTCTGT GGCTACTTAC AGCTATTTAA TGTTGTATTT GTTCAATTA TTTGGCAAAT	3420
TATAACAAAA AGACAACAAA ACCAACGTTG GTTTTGTTGT CTTTCTTTT TTAGATTGAA	3480
CGTAATTGGG CAATCCGTTT TTTGACAGCC TCTTGTTTTT CAAGGTAGTC CTTTCTTTT	3540
GCACGTTCTG CTTCCACGAC CTCATCTGGT GCATTTGACA CAAAACGTTT ATTTGAAAGT	3600
TTGCCTTGCA CACGTTTGAC TTCTTTTGTC CATTTGTCTA ATTCTTTTTC TAAGCGAGCA	3660
ATTTCTTCCT CGATATTAAT CAACCCAGCT AATGGTAGGA ATAGTTCTGC GCCTGTAAA	3720
ACAGCGGACA TGGCTAATTC TGGCGCTTCT ATTTACGAC TGATGACTAA TTCCTCTGGG	3780
TTACAGAAGC GTTCCAAGTA GCTAGTATTA GCTGTAAAAA ATTCTTCTAC TTCTGTATCG	3840
TTTGTTTTAA TTAACAAGGT AATTGGTTTA GAAAGTGGTG TATTCACCTC TGCTCGAATA	3900
TTCCGAACCG AACGAATAAC TTCTTTCAGA ACTTCCATTC CTCGAGCTGC TGCTTCATCG	3960
TTAAATTCTT CATGAACAAC TGGGTACTCT GCCACAACCA ATGACTCACC TTGATGCGGA	4020
ATTTTTTCCC AAATTTCTTC TGTTACAAAC GGCATAATTG GATGCAATAA GCGTAAGATT	4080
TGATCCAATG TGTACACCAA GATGCTCCGT GTTGTTTGTT TAGCTGCTTC ATTGTCTCCG	4140
TAAAGAATTT CTTTACTTAT TTCAATATAC CAATCACAGA AATCATCCCA GATGAAGTTG	4200
TATAATTGGC GACCTGCTTC ACCGAATTCA AAGCGATCAA ATAATTCTGT CACACGGGCG	4260
ACTGTTTCGT TTAAGCGCGT TAAATCCAA CGATCAGCGA CAGTCTTTTC ACCACTAAAG	4320
TCAATATCTG CAGCAGTCAT ACCTTCTACA TTCATGATAA CGAAACGACT TGCATTCCAG	4380
ATCTTATTGA TAAAGTTCCA AGAAGCATCC ATTTTTCAT AACTAAAACG AACATCTTGA	4440
CCAGGTGCTG AACCATTTGA TAAGAACCAA CGTAACGCAT CTGCGCCATA TTTTTCGATG	4500
ACATCCATTG GATCAATCCC GTTCCCAAT GATTTACTCA TTTTGCGTCC TTGCTCATCA	4560
CGAATTAAGC CGTGAATTAA GACATTTTGG AAGGGACGTT CACCTGTAAA CTCTAAGCTT	4620
TGGAATATCA TGCGACTTAC CCAGAAGAAG ATGATGTCAT AGCCAGTAAC TAATGTGCTT	4680
GTTGGGAAGT AACGTTGGTA ATCTTCGCTT GCTTCATTCG GCCAACCCAT CGTTGAAAT	4740
GGCCATAAAG CAGAACTAAA CCATGTATCT AAGACGTCAC TATCTTGAAC CCAGTTTTCA	4800
CTGTGCGCTG GCTCTTCCAT CCCAACATAC ATCTCGCCCG TTTCTTTGTG ATACCAAGCA	4860
GGAATTTGGT GGCCCCACCA TAATTGTCGC GAAATAACCC AGTCATGTAC ATTTTCCATC	4920
CAACGTAAAA ATGTTTGATT GAAGCGTGGT GGATAAAATT CAACAGCGTC TTCTGTTTCT	4980
TGATTTTTC TGGCTTTTTC AGCTAATGGT CCCATTTTTA CAAACCATTG GGTGATAAA	5040

CGAGGTTCCA CGACAACGCC TGTTCTGTTCT GAATGGCCCA CGCTGTGATT CATTGTTTCG	5100
ATCTTAATTA AGCGACCTAG TTCTTTCAAG TCAGAGACAA TCGCTTTACG TGCAGCAAAA	5160
CGATCCATTs CCTcGTATTT TCCAGCCAGT TCATTcATcG TGCCGTCTTC GTTCATAACA	5220
TTTACTCGAG GTAAGTCATG ACGATTACCG ACTTCAAAGT CATTGGATC ATGGGCAGGT	5280
GTGATTTTTTA CAACCCCTGT TCCAAATTCC ATATCTACAT AGTCATCTGC AATAATTGGA	5340
ATTTCTTTAT CGACTAATGG CAAGACAACGT GTCTTACCAA TAAGTTCTTG ATACCGTTCA	5400
TCTTCTGGAT GGACCGCAAT CGCCGTATCT CCCAACATTG TTTCAGGACG CGTTGTCGCA	5460
ATTTCAACAA CGCCAGAACC ATCTGATAAT GGATAGCTCA TGTGGTAAAA AGCTCCTTCA	5520
ATATCTTTAT GGATTACTTC AATATCAGAC AATGCTGTTT TCGCTTTTGG ATCCCAGTTA	5580
ATGATATATT CACCACGGTA AATCAAATCT TTCTCGTAAA GAGAGACAAA CACTTTACGA	5640
ACGGCCTCAG ATAGCCCTTC ATCTAAAGTA AAACGTTTAC GGCTGTAATC TAAAGAAAGA	5700
CCCATTTTTG CCCACTGTTT ACGAATGTGA GAAGCATATT CTTCTTTCCA TTCCCACACT	5760
TGATCAACAA ATTTTTCACG ACCTAAGTCG TAACGTGAAA TTCCTTGTTG TGCTAGTTTC	5820
TCTTCTACTT TTGCTTGCGT GGCAATTCGG GCATGGTCCA TTCCTGGCAG CCATAACGTA	5880
TCAAAGCCTT GCATTCTTTT TTGGCGGATA ATCATATCTT GTAATGTTGT ATCCCAAGCA	5940
TGCCCTAAGT GTAGTTTCCC TGTTACATTG GGTGGTGGAA TAACAATTGA ATAGGGctTC	6000
GCTTTTTTGT CGCCACTAGG TTAAATAAAA TCTTGATCTA GCCATTTTGT ATAACGGCCA	6060
GCTTCGATTT CTGTCGGTTG ATATTTCTGT GGTAAGTTTT TTTCTTCTGA CATTTTTTCC	6120
CCTTCTTTCT TAAATACGC GATTTGGAGC GGTAATTATT TTACAAATGA GTTTTTTAGG	6180
CTAAACACTA AAATACGCCC TAAATACTT TCGTATTTTA GGgCsTAGaT TCATCTTATT	6240
TACGCGGTAC CACCTAAAt GCATTtGAAT tGAATtCCTT CAAATGCCAC TCATtCGTGA	6300
GATAACGTTT ACGAAACGAA TCGTTCTACT AACGTACTGT GTTCAAACCA TTGACTCACA	6360
AGCTACCTTC TCACTGGTTG TGTAAGAGTC TTGCACCTGA ACGACTTTTT CTCTAAAACG	6420
ACGCCAATGA TACTCCTCTT GTTCGCTGTC TTTGCTTCAT TCTACAATGA AGCGCAAgcA	6480
AAGTCAATTT GTAATCTAGA GGGATTACAC TTTTTTTGCA GATTCTAACG CTTGTTGATA	6540
ATCGGGTTCC GACGAAACTT CTGAGACGAT TTCTTCATAA ACTAACGTTT CTTCTGGGTC	6600
AATCACAAAA ATAGCACGTG CTAAACGGCC CATTCTGGA ATATACAAAC CATAAGCTGC	6660
ACCAAATGAG TCTTCAGTAT CATGAAGCAT TTCCATTTTCG ACACCCTCAG CGGCACACCA	6720
ATTTGCTTGT TCTTCAACTG TGTTATTGGA AATCGTGATA ATCTGAACAC CGTCTAATTT	6780
CGCCGCTTCT TGATTGAAAC GTTTCGTTTG CAGTGAACAA ACACGTGTGT CGATATCAGG	6840
AACCACACTA ATTAAGAACG TTTTACCTTT ATAATCTGCC AAGTTGATTT CTTGGTTGTT	6900
TAAATTTTTC AAAGAAAAGA CGGGAGCTTT TGTGCCAACT TTAGGCTGCT CACCTGTCAA	6960
TTCTAATACG TGCCCTTTTC TTGTAACATT CATTCTGAAA TCCTCCTTTT CCACTTTCAA	7020

TAATAAGTAT AGTAGAGGAT TTTTATGAAG ACAATTTTAT TGCTTTGGGC TATTTGCTAG	7080
TGAAAAAGTT TCATTAAAT AAATCAGTGT TTGCAGTTCC GTTGTTAAAT CAATATATTG	7140
CACACTAACG TTAGCTGGCA CACTGACACG GTCTGGCGCA AAATTTAAAA TCGCAGTCAC	7200
ACCCGCTTGT ACAATTTTGT CGATGGCTTT TTGAGCATGG TGAATTGGTA CTGTAGAAAT	7260
GGCTACCGTG ACCCCTTCTT GTCGAACGAA TGCTTCTAAC TCAGACATAT CATGAACTAA	7320
TAATCCATTA ATCGTGGTGC CAACTAAAGC CGAATCATTA TCAAAAGCAC AGACAATATT	7380
CAAATTCTCG TTGCGTCTAA AATTATTTTT TAATAATGCC TTGCCTAAAT TGCCACAACC	7440
AATCAGAGCA ATTCTTTTTT CTTCTTGCCT ATTTAAATA TTAAGAATA CCTCAATTAA	7500
ATACgGCACA TCGTn	7515

## (2) INFORMATION FOR SEQ ID NO: 29:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7156 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

ATATGGCCTT CACTTTAGTG AAGAGATTTA ATTGTCTTTA CAGATAAACG CTTAATTTTA	60
GTCGATAAAC AAGGAATTAC AGGTAAAAAA GTTGATTATA AATCGATTCC CTACAAATCT	120
ATTTACGCT TTTCTGTTGA AACTAGTGA CATTTCGACC TAGATGCAGA ACTAAAGATA	180
TGGATTTCAA GTACTGAGTT ACCTTCAGTA AGTCTTCAAT TTAGAAAAGA TAAAGATATT	240
GTTGCTATTC AACAAGCTTT AGCAGCTGCT GTATTATCTT GAATACACAA AAAAAGACAG	300
GAAAGCGACC AACTTTCCTG TCTTTTAAAT TktAGATAAA ATAAAAAAGG CTTCTCAGCC	360
TTGATACGTT CTTATTCAA TCAGACACAT GGCGGCACTT GCTTAGTGCT GCTTCCTTCC	420
GGATCTGACA CAATTCAC TA GCGGCCATT GTCCTAGCCT TTCGGCAAAA CCTAGTATAA	480
CGTATTTTTA TTCTTTTGAC TAGTCTTTCT TTTGATTTTT TAAAAGTTTC TCTTCTTTTT	540
TCTTTGCCCC TAACTTTTTA AAGAACTGAA CTAGCAATAA GCGACACTCT TCTTCTAAAA	600
CGCCTGCTTC GACATAGGCC ACATGATTGA AGCGTTCATC TTCTAATAAA TTCATCAATG	660
TGCCAGCAGT GCCTGCTTTA GGATCAGTTG CGCCATAATA AACTTCTGGA ATTCGTGCCA	720
AAAGCATCGC CCCACTACAC ATTGGACAAG GCTCTAAAGT TACAAATAAT TGGGTTTCTT	780
CTAGACGCCA ATTTTCTATC CCGCGACACG CTTCTTGGAT TGCATACATT TCTGCATGCG	840
CCGTTGCTTC TTGCCGTGCT TCTCTCAGAT TATGCCCGCG GCCAATAATT TCCCCTTGCC	900
GAACAATTAC AGCTCCAATA GGCACCTCAG CCAAACCTTC TGCTTTTTTC GCTTCTGCAA	960
TGGCTTCTCG CATAAAAAAT TCTTTTTCTT CTTGTGTTAA CGTTGTTTCT TTTTTCCCA	1020
ATTAATTTC ACCTTTCGGC TTTTCCAC TAAACTTCTG TCTTGCTTCA TGGTACAATA	1080



AATCAGTGAG GTGTGTCTAT GATTATTGAA AATTTATCTC AGCTTTATCC CAAAGGCTAT	1140
TTATCAAAAA CAATTGGTTC CGAAGCCACC TATTCAATTC CTGTTGACGA CCAATTTTTC	1200
ATTGTGAATA AAGCTTTCTT AAGTATTAAA GAACAGCATT TGTGGAAGC GCTTTTCCCA	1260
ATTTCTGAAA aCCCAACAAT TACTGGTAAC CATCCTTGGT TCAGTTACTT ATTTCAACAG	1320
GCAAAATTAC CTGCCGAAGG GACCTTTCGT ATGTTGCAAA TTCAAACGAA TGTCACAAAA	1380
GAATTACAAG CTGAATGGCA ATTTAATTTA ACAAAGATGT TTCCAGATAC AGTTGATTGT	1440
TTCAGTCCAA GTAATAATAT GTATATTTTA GTTGAAGAAC AGTCAAAAAA TACCTTTCAA	1500
CAAGAAGAAA TTCAAGGAT CTTTTTAACA TTAGATACAG ACTTTGATTG CACTAGTGCT	1560
GTATTTGTTG GCAATTTTTA TTCCTCTGAA GACATTCTCC GCGGTTGTTT CCATGAAGAA	1620
CAACGTATTT TTTCAGAAGA ACTTAATTCC TCTAGTCGGA CGACTGTTT TAGTTTAACT	1680
GATGTTGCCC TCCACTATTT TACCAAGGAA GCCATGAGCC AAAATGTACT TGTTGAGTAT	1740
TATCGTCGTT TATTAAATAA AGACACCGAT ATTCAGCCAA TCATCAGCGC TTTATGGAAA	1800
AATCAAGGCA ATATCAGCTC CACTGCAAAA GATTTATTTA TGCATCGAAA TACTTTACAC	1860
TATCGGTTGG AAAAGTTTTT TGAACAACT GGCTTATCTT TAAAAAAAT GGATGATTTA	1920
ATCTTTTGTT ATTTGTTATT AAGAAAGTAG TACGTCCATT TTTTACTCGT ATGTTTTTTA	1980
TTCCCAATTA GAAAGTAGGC GTTCCAATGA AAATTACCGC ATATGTCGCT AGTGCATTTA	2040
GTAAAAACCA CGAAGGTGGC AACAAAGCAG GCGTTGTTTT TATGACACAG CCCTTAACGA	2100
GAACCCAAAA AATGGCCATC GCCAAAGAAC TAGGATTTGC AGAACTGCT TTTATTTTCA	2160
ACTCAAAGAT TGCTGATTTT CGCCTTGAAT ATTTTACACC AAAAGAAGAA GTCGACTTAT	2220
GTGGCCACGC GACAATTGGC GCTTTCACCA TTATGAAGCA CTAAATAAA TTATTGCAAA	2280
CAGCATATAC AATCGAAGC AACAGCGGTT TGCTCGCAAT CACGGCTAAC GGTGAACAAC	2340
TATTTATGGA ACAAAGTGC CCAACTTTTT ATGATACGCT CACTTTAGAA GAATTGAGCG	2400
GCTGCTTTGA TTAAACGCA GTGCAACTG CTTTCTCTAT CCAGATTGTG TCAACGGGCC	2460
TAAAGATAT TTTGGTTCCC ATTAAAAATG CGCAATTATT ACAGGATTTA CAACCAAAT	2520
TTGAAAAAT TAAGTTACTT AGTCAAAAGT ATGGAGTTAT TGGGATGCAT CTTTTTGCTT	2580
TGTCTGAAGA AAATATCATT TGTAGGAATT TTGCACCCCT TTATGATATC GATGAAGAAG	2640
CCGCTACTGG TACATCAAAT GGTGCGTTAg CTTGTTATCT GCACCATCAC CAAATTTTAC	2700
AAAAAACCAA TTACACCTTT GCCCAAGGTT ATACCTTAAA CGCTCCTTCA GAAGTCTTAG	2760
TGCGTTTGAA CACTACATCC CAAAAGACAA TTGATCAAGT TCTAGTTGGT GGCAGAGGCT	2820
ATTATGTTGA AACAAAAGAA TTATTTTAT AAAAAAAGA TGGCCTAGGC CAcTTTTTTa	2880
ACTTCaTTGC ATCTTCAATA AATGTGAAAA AGCGTTCCTT TAAAGAATCT GTATTGTCTT	2940
TAACTTTTTT GCCATATTCA GCTAATTTTG AACCGCCATC TTTACACGA TCCACAACAT	3000
TTAAACAGA ATCTAATTCA TCATCGGATA AATCATCGAC AATCGATAAT AATTTTGGT	3060

TTCCATCAAA TTTATCGTCT ACAAACCTTT TAACTTTATA ACGATTGGAA ACATGAGATA	3120
CCTTCTTAAT AATTTTTTCA GAAGCGATGA CTGCAACAGA GACACTTGCa ACTGCAGcAA	3180
TGCTTAAACC GATACCAATT TTTGTTGATG CTTTCaTTTC CTTCAGCTCC TCTGTTATTT	3240
TCTACTTTAA TCATAACATA TTATTACTTT wAkGCGTAwT rGAAaGrCCT TTACTAGCAC	3300
TTAGTCTCGT CGGTTATTGC CACCAAATAA AATAACCAAA CGTAAAAGTT GTAAAAAGGT	3360
CGAAAGCGCC GCTGCTACGT AGGTAAGTGC AGCTGCAAAA AGAACTTTTC TGGCCATTGG	3420
GACTTCTTCT TCTGTTAATA ACCCACCTTC AGAAAGAATA CTTAATGCAC GTCTAGACGC	3480
ATTAACTCC ACCGGTAAAG TCACCAATTG AAAAATCAAG GCTAATGAAA AAGCCAAAAT	3540
TCCAATGTTG ATTAATGTTT GATTCCAAC TAAATAAACA CCAAGCAAAA TTAGTGAAAA	3600
GGAAATGGTT GAACCGATGT TGGCAACTGG AACAATAGCT GCACGAAGAC GTAAAGGCAC	3660
ATAGTTCGTA TGATCTTGTA CCGCATGTCC GCATTCATGA GCCGCCACGC CAATAGCCGC	3720
GACCGATGTC GATTGAGCAG TGGCTTCGGA TAAGCTCAAC ATTTTGTTTC CAGAATTATA	3780
ATTGTCGGTT AAATCACCCG CAATTGCTG AACTCCAACA TCATTAATTT GTTCTTTTTC	3840
TAAAATATAT TGTGCTGCCT GCGTGCCTGT AACATGTTTA CTGCTTCGTA CTTGATCATA	3900
TTTACGAAAG GTACTGTAA CATATGCGGA GGCCGCACCA GAAATCGCAA TTCCGATAAT	3960
TACCAATATA TAGGTTGGAT CGATTCCAAA ATAAAAGGGC ATCATTGTCC TCTACCTCCA	4020
TCTTTTTTTC TTTATTGTAG CATGATAGTG AAAAAAGGG AGCGACAAAC GTTGTCTTTT	4080
GTTTAAAAAA ATGATGAATT TTTCAAGAAA GTGCCGAAAC TAACAAAATT GTCACTCAAT	4140
TAACGAAAAA AAATATGGCA AATTTCCTAA AAAAAAGGCA TACTATTCCT TGAGGTGAGG	4200
AAAATGAAAA CATTACTAAA AATATTAATT CCCATAGTCA TTTTAGCCGG TGGTAGCTGG	4260
GGCGTTTATA ATTACTATTA CGGTGGAGAA GCCTATTATA CACAAATCAC AACATCTGGT	4320
GAAAAGAAAG ATGAAAAAAC AAATTCTGGT GAAGCAATGA CCATTTATTA CTATCAACAA	4380
CCTGCGTTTA ATAAAAACGG TGAAGAAAA ACCGTGGAAT TAAACGAATC CCGTGATCAA	4440
CCTTTGCGTA TGAAAGCCTA TCTAAAATTG AAAGTTAATC CCCGTAAAGG CGTGATCAGC	4500
TGGAATGAAG TGACAGAAAA AGAAGTCCCT GAAAAAGCTT TAGAAAAATT AAAATAAAGA	4560
AAAAACCGTT GGCAAGCGAC CAACGGTTTT TTCTTTATAA TTGATTACTA ATTTTTCGA	4620
TATTTTTTAG CATAATGGAC ACTGTGCCAT CTGGATTATT AACAAATTC AATTAAATCGC	4680
CATTGCGGTA AACATCTAAC GGAACAATTA ATTCAATACC GTTATCCATT TTTAACTTCT	4740
GTTTGCCAAA TTTCTTTTCA GAAATTTTCG GAGATTCTCT GACTGGCGGT GTTTTGTGCA	4800
CAAACTTCGA TTCATTCCT TCTTCTGGT AAGCTAATTT AGCTGATACA TTTTCTTTAA	4860
ACACTTGCTC AGCAATCTGC TCATTGTCAA TCGTCCCCGT TTCCTCAATG CTATCATACA	4920
CCGcTTCTTT TACAGAAGCC ATTAGTTCAA ACTCATCTTC GTTAAATTT TTTCCAATGC	4980
GTTTGACTGC CTTTTTAATT TCTTTGATAT TTTCTCCAT GGAAGGTGCC GCTTGCCTCT	5040

CAATCACTTT TTCAGAGAAA TACCAGATTT TTTCTCCTGA AAATTCATAT CTTTCTCTA	5100
GTAATTCAAA CGCTAATGTA TCTAAGTCAA CCGTAATCCC TTCACTGGT TTTGTGATT	5160
TTGAAGACAA AATTGCTCGG TTGATAATCA GTTTGTTATA AACGGATTCT TCTTCATAGT	5220
CGACGTAATG CGTGTAACCT TCTTTATAGT TTAATTTAAT CATTGCGACA TGTAACACCG	5280
TATCTAATTC ATATAAAATG AACAAAATAT CAGCACTAGG CGCATCTTCG CTTCTGAAT	5340
AAACATCATA CCAATGTTGG GTAACCTCTT GGCTTTTCTT TATAAAATCA CTAGGAATCC	5400
CACGTAACGT GTCAACAAAG GAACTATTTT CCGCCAATGT TCCAGTCTTT GTTGTGCAG	5460
TCGAAAGTTT TGAAATTTTT GCAGTTAAAT ACGTTCGAAT ATACTCGGTC GTTAAATCTA	5520
GTTCCTTTTC AGAGAAAACC GGCGTCCCAG TTTCTCGATC AATAATGTGT AAAATTGCTT	5580
TTTTTAAATA AATATCCATT TTTACGCTCT CTTTTTCTCA ACCTTTTGTAG TCTAACTGAA	5640
AGAGAGAAAA AAGCCAATAG TTTCTATAAA AAGAAGAAGA ATTCGTAGAC AAATGTCTAC	5700
GAATTCTTCT TTAATTCAAA AAGCGATTAC CCTTCTTCGC TGTATAACAT CCCTGTTGCT	5760
TCATCAACTG GATCGACTTT ATTAAAAATG CCTTCTTTC TGAAATATA AGTTAAGCCC	5820
ATTGGTACGA TAATGGCAAT CACCATTGCG CCTAAAAACG GCAAGTAATA TTGTGGCACA	5880
ATAGATAAAA TTCCTGGCAA GCCACCGACG CCGATACTGG TCGCACGAAC GCCCATTA	5940
GTGGCAAACA TTCCTGCCAT TGCTGAACCA ATCATGCCCG CAACAAATGG ATAAACATAT	6000
TTTAAGTTGA TCCCAAACAT CGCTGGTTCA GTCACACCTA GGTAACAAGA AATCATCGCT	6060
GGAATCGAAA CTTGTCTCTC TTTTtGTTT CCGCGATGCA TAAAGaTAAC CGCTAAAACA	6120
GCTGAACCTT GTGCAATATT TGATAGAGCA ATCATTGGCC ATAAGTTGGT TGACTGAAAG	6180
TCAGCAATCA ACTGTAAATC AATCGCATTA CTCATATGAT GCAACCCAGT GATGACTAAT	6240
GGTGGGTACA TAAAGCCAAA GACGGCGCCA AATAACCAAT TAAATGTTGA CGTTAATCCT	6300
GCATTTACGA TATCAGAAAC CCAACTACCA ATCGTCCAAC CGATTGGTCC TAAAATGACA	6360
TGTGCAGCCA ATACTGTTGG TACTAAAGCA AACAATGGGA CAAAAATCAT TGAAATTGCT	6420
TGAGGAATCA CTTTTCTAAA GAAAATTTCA AGATAGGCAA GCAAGAAGCC GGCTAACATG	6480
GCTGGGATAA CTTGCGCTTG ATAACCAATC ATGTTAACTT GAGCAAACCC AAAATCCCAG	6540
AAAGGAATTT CTGCTGCTTT AGTCGTCGCA ACTGAATAAG CATTGAGAAG TTGTGGTGAG	6600
ACTAATGTAA TCCCTAGGAC AATCCCCAAG ATTTGTGTGG TTCCCATTTT CTTCTGTAATA	6660
CTCCACGTAA TCCCAACTGG CAAGAAGTGG AAAATTGTTT GCCGATTAAAC CATAAGAATG	6720
AATTGACTCC ATTCCAAAT TGAGAACTT CCACGATTGT TTGACCATCA AGAAAGCCAA	6780
ATGGAACCCC TTCCAAAATA TTGCGGAACC CTAAATCAA TCCCCGATA ACAAGTGCAG	6840
GAATCAATGG TGTAATAATC TCAGCTAAAA CAGAAACAAC TCGTTGAACA GGAATCAAT	6900
TACTTTTGGC AGCTGCTTTC CCTTCTTCTT TTGAAACCCC TTCTAAACCG GAAATTGCAG	6960
AAAAATCATT ATAAAATTGT GGGACTTCAT TCCCAATAAT TACTTGAAAC TGTCCAGCAT	7020

TTGTAAAGGT TCCTTTTACA GAAGGAATAT CTTCAATCAC AGGTACATCC GCTTTTTTTTG	7080
GATCTTTTAA TACAAAACGC ATTCGAGTTG CACAATGAGA AACAGCTGAA ATATTTTCTT	7140
TGCCTCCTAC TGCTTT	7156

## (2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1282 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

ATTCGTGTA ATTCCTAACTG GAATTCCTAA TACGAACCTA GAAACGTCGT ATCCATTTCT	60
TGCTTTAATG TCAGTTGTTT TTGGACCAGT AGCTGGAGGG CTAATTGGTT TGATCgGTCA	120
CACgTTAAAA GATTTTACCA CGTATGGTAG CGCTTGGTGG AGCTGGATTA TTTGTTCAAGG	180
AATTATTGGT ATTATTTTGG GCTTCGCTGG TCGTAAAATG GATCTTCAAC ATGGTGAATT	240
CACAACAAAT GACATGGTGC GTTTCATAT TTTCCAAGCC TTTGGCAATA TCGTCGTTTG	300
GGGCTTAATC GCGCCTAGCT TGGATATTTT GATTTATAGT GAACCCGCAA GTAAAGTGTT	360
TACACAAGGC GTATTCGCGA CTGTTTCAAA TATTGTCGCT GTGGGGATTA TTGGAACATT	420
GTTGATGAAA GCTTATGCAT CAACAAGAAC AAAGAAAGGC AGTTTATCAA AAGATTAAGA	480
AACTGTCGAA ATAATTTAAA TGGTTAAGGG TCTGCCCTAG TAATTACTTT GCTAATGCTC	540
AGACCCTTTC TTTTATGGAG AAAGTGAAGT GAGGCCGAAT GAAAAACCA ATCATTACTT	600
TTAATAATTT TTCCTTTCAA TATCATAGCC AATCTGAACC AACATTAAAG GGGATTCAAT	660
TAACCATCTA TGAAGGAGAA AAGGTTTTAA TCGTTGGGCC TAGTGGTAGT GGTAAATCAA	720
CATTGGCGCA ATGTATCAAC GGGTTAATTC CCAATATTTA TGAAGGTGAG ATTCAAGGAA	780
CAGCTACGGT TGCTGGTAAA AATATCCAAG AAACAAGTCT ATTTGATTTG TCTTTTGATG	840
TAGGAAGTGT TTTACAAGAT ACCGATGGGC AGTTTATCGG TTTAACGGTT GCTGAGGATA	900
TTGCTTTTGC TTTAGAAAAC GATGCAGTTG AACAAGCAGA AATGAAAAAA GCGGTTCAAG	960
AATGGTCCGA GATTGTTGAG CTGAATCAGT TATTGCAACA CCGTCCACAA GATTTGTCAG	1020
GTGGTCAAAA ACAACGTGTA TCTATGGCGG GCGTTCTAAT TAACCAATCA AAAATTTTAC	1080
TTTTTGATGA ACCTTTGGCT AATCTTGACC CACGAGCGGG ACAAGAAACA ATGACTTTGA	1140
TTGACACCAT TCAACAAGAA ACAAAGCGA CTGTTTTAAT AATAGrACaT CGCTTAGAGG	1200
ACGtCCTTTG CGAATCAGTT GATCGAATTA TTGTCATGAA TGAgGGCACa ATTATTcTGA	1260
TACAACGCCA GATGAATTGT TA	1282

## (2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5840 base pairs
  - (B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TAAGGATTGT TACTGCTTTG CAGGCAAAAC CAGAGTCTTC CATTAACACA ATCAGTGGAG	60
GGATTCTGGG CTTTTTTTAT AAAAAAGTGA GGTTCATTC ATCATGAGGA TTGAAAAAGT	120
TTTAAATAAT AATGTCGTGC TTTCAAGAAA TGAGCAAGGC GAAGAAATG TTTATATGGG	180
ACGCGGCTTA GYCTTTCAAA AGAAGATTGG AGACACGATC AATCCAGATT ATGTTGAAAA	240
GGAATTCGTG TTAAAGGACT CCGCTATGGC AGGGCAATTC CAGCAGCTGT TTGAAGACGT	300
CCCCACCAAA GAAGTAGAAG TCGTCAAACA AATCGTTGAT TTAGCAGAAA CAACGCTAGC	360
GATTGAGCTT TCATCAAATA TTTATTTAAC TTTAACGGAT CATATTCACT ATGCGTTACT	420
TCGAGCCAAA GAAGGAATG ATATTCCGAA TCCTCTTGTA TTCGAAACAC GCAAATTTTA	480
TCCTAAAGAA TTTGAAATCG CAAAGCAAGC ATTAGCGATT ATTGCTGAAA AGCTAGGGGT	540
TCAATTTTCA GAAATGAAG CGGGCTTCAT TGCTTTTCAT ATTGTGAATG CTGAACAAGG	600
AAATGGCAAT ATGGAAGTTA CCATGGAAGC AACCAAGATG GTTCGTGATA TTTAACGAT	660
TATTAGTCGC TACTTTGGAC AGGTGTTTGA TGAAGATTCA TTGAACTATC AACGAATTGT	720
TACACATCTT CaATATTTTG CGCAACGTTA TTTAAACAA GAAGCGCATG aTGAGGAGGA	780
CGAATTTCTT TtGcAtGg TCCAAGGAAA GTATCCTAAG GCGTTTCAAG CGGTCCAACG	840
TATTAATGAG TATTTATTAA AAAGCTATGA TCGACCAATT GATCAAGCGG AAATGATTTA	900
TTTAACCATT CATATCCAAA GAGTGGTCAA TGATAAAAAA GCATAACAAC CAATAATCGC	960
GGATTGTTAC TGATAATGCA GGCAAAACCT AGATTGAAAA AACCGATAAC GATTTTCGCA	1020
GACTGCGCAC TGTGAAGAGA AGGATCTGGC TCTTTTTCAA TGTAGGTTTT TTTGATATTA	1080
AAAATTAATG GAAATGAGGt CTAGAGATGA GTAAAAATCA AGAAATTGCA GCACGTGTGT	1140
TAAAAGCTGT TGGGGGCGAA GsCamCGTTA ATAGkGTTGk TCACTGTGCy ACACGCTTAC	1200
GtTTTAAATT AAAaGrTGAA AATAAAGCAG ATACAGCTGC ATTGAATGCT GACCCAGATG	1260
TTATTCAGGT AGTTCAAAGT GCAGGACAAT ATCAAGTGGT TATTGGCAGT CACGTTAGTa	1320
TGTCTATAAG GATTTGATGG CCAACAGTGG ACTAGGTAAT GATTCTGATA ATAGAGAAAA	1380
AGAAAGTGCC GGAAATATCT TTAACCGTTT AATTGATATT ATTTCTTCTA TTTTACACC	1440
GTTTTTAGGT GCAATGGCCG GCGCTGGGGT TCTAAAAGGC TTCTTAACAT TGGCTGTAAC	1500
TATGGGCTGG TTAGCTGACA CTTCAGGCGT ATATCTTGTT CTCTTTTCGA TTGCAGATGG	1560
TTTATTTACT TTCTTACCAA TTATGTTAGC GTTTACGGCT GCTAAAAAAT TCAATACTAA	1620
TCCCTTTTAA GCGGTTGCTT TGGCGATGGC CTTGGTTCAT CCCAGCATTa CTGCTTTAGC	1680
AGGCAAAACA ATTAGCTTTG CTGGTCTTCC AGTGATTATT GGGCCAAGTG GCTATACTTC	1740
TTCGGTTTTA CCAATTATTT nGGCTGtTTT TGCACAAAGT TATGTAGAAC GTTCTTCAA	1800

AAAAGTAATT CCGAGTTTTC TACAAATTAT CTGCGTTCCG TTAGCTGTTT TCCTTATTAT	1860
GGCACCTGTT ACCTTCTTAG CTATTGGTCC TATCGGGACG GTCATCGGCG ATTGGTTGGG	1920
ACAAGGATAC AACGCAATTT ACGCTTTTAG TCCAATTATC GCAGGGTTAT TAATGGGTTC	1980
GTTGTGGCAA GTCTTGGTAA TGTTTGGTAT GCATTGGGGC TTTGTACCAA TTATGATGTT	2040
AAACTTAACA CAAGGTGGCG ATACGATGGT ACCGATGTTA TTACCAGCCG TTATTGCACA	2100
AGGCGGGGCT GCTTTAGCTG TCTTTTTCCT AACAAAAAAT GTGAACTAA AAGGTTTGGC	2160
TTTGTCTTCA AGTATTACGA CTATTTTGG AATTACTGAA CCAACTGTAT ATGGCGTGAC	2220
TTTACCATTG AAAAAACCAT TTATTGCAGC TTGTATTGGT GGCGGTATCG GTGGTGCATT	2280
TGTGGCTATG AATCACGTGA AAAACTTTAC GTTTGGCTTG GTTAGTATGT TGAGCTTGCC	2340
TGGCTTTATT CCTGCAGAGA CAAAAGATAC TGCACCGATG ATTACTGGTG CAATTGGTGC	2400
CGGAATTGCC TTTATCATTG CGTTGTCTT AACGTTTGTC TTACGTTTGG AAGATCAACC	2460
TAATCCAGAA ACAGCAACTG AAAAAACAGA AACTGATAAG ATGGTGGCAC CTGTAAAAAC	2520
GAATCAAGAA GACAAAATTA TTTTAGCAAG TCCACTTCAA GGTGAAATTT TACCGCTAGA	2580
AAAAGTACAA GACCCTGTTT TTGCTTCAGG TGCTTTAGGA AAAGGTGTTG CAATTGAGCC	2640
GACTGAAGGC AACTGTATG CCCCCGAGA TGGTGAAATC ACCACATTAT TTCCGACAGG	2700
ACATGCTGTT GGCTTGACGA CAACAGAGGG CGTTGAATTA TTAATGCATA TTGGCATGGA	2760
TACGGTCGAA TTAGATGGTA AAGGCTTGA ATTATCAGTG AAACAAGGTG ATTCTGTAA	2820
AAAAGGAGAT TTGCTAGTTA CTTTGTATAT TGCTGCCATT AAAGAAGCTG GTTATCCGGT	2880
AGTTACACCG ATTGTGGTAA CGAATACGAA TGATTATTTA GATGTTTTGG ATATGAACCA	2940
AACAGACGTG TTACATGGTG AAGACTTTTT AGCCATTATT AAATAAAAGA CAATTGGAGG	3000
AGGAACAAGT ATGACAACAA CGACAAAATT TCCAAAAGGA TTTTATGGG GCGGTGCGAC	3060
CGCAGCCAAT CAATTAGAAG GGGCTTACTT AGCCGATGGT AAAGGGTTAT CTGTGGCAGA	3120
TGCGATGCCG GGAGGAAAAC AACGTTTCGC TATTTTAGGA GATGAAGCGT TTGATTGGAC	3180
GATTGmCGAA ACGAAATATC GCTACCCTAA TCATACAGGA ATTGATCACT ATGATCGCTT	3240
TAAAGAGGAC ATTGCGTTGT TTGCTAAAAT GGGCTTCAAA TGTTATCGCT TTTCAATTGC	3300
GTGGTCGCGT ATTTTTCAC AAGGAGATGA AACACAACCG AATGAAGCCG GCTTGAAATT	3360
TTATGATGCG GTTATCGATG AATGTTTAAG CTATAACATT GAACCAGTAA TTACGATTTT	3420
TCATTATGAA ATGCCACTGC ATTTAGCGAA AGAATACGGC GGGTGAAAA ATCGTCAATT	3480
AGTTGACTTT TATGAACGTT TTGCGGAAAC TGTTTAAACA CGTTACCACC AAAAAGTTAA	3540
CTATTGGATG ACATTCAATG AAATTAATTC AGCTTTTCAT TTCCCAGCGC TTAGTCAAGG	3600
ATTGGTAAAA AGCAATGGAG CCAATGATTA TCAAAACATT TTCCAAGCAT GGCATAATCA	3660
ATTTGTTGCA AGTAGTAAAG CTGTAAAAAT TGGCCATGAA TTAAATCCAG AATTACAAAT	3720
TGGATGTATG ATTATTTATG CCACAACGTA TGGCATTGAT TCGAATCCAG TCAATCAAGT	3780

TGCTACAATG ATTGAAAACC AAGAGTTTAA TTATTATTGT ACAGATGTTT AAGTTCGTGG	3840
TGAATATCCA GCATACGCAG AaCGCATGTA TCAAAAATAT GCTGTGAAAG ACTTAGTTAT	3900
TGAAGAAGGA GATTTAGAAT TATTAAAAGA ATACCCAGTG GATTACATTG GCTTTAGTTA	3960
TTATATGTCG ACTGCTGTCG ATGTTACTGG TACTACCAAC GATACAGCAA ATGGGAATCT	4020
TTTAGGCGGC GTTAAAAATC CCTTCTTAGA AGCAAGTGAA TGGGGTTGGC AAATTGATCC	4080
CGAAGGCTTA CGAATCGCCC TAAATGAATT ATATAATCGT TATCAAAAAC CTTTATTTAT	4140
TGTGGAAGAC GGCTTAGGTG CAATTGATAA AGTAGACGAA AATTTCTATG TGGCAGACGA	4200
CTATCGGATT GATTATTTAC GTCGTCATAT TGAAGCAATG GCAGAAGCAG TTGCAGACGG	4260
CGTCGATTTA ATGGGCTATA CGCCGTGGGG CTGTATCGAT TTAGTCAGTG CTCCACAGG	4320
TGAAATGAGT AAACGGTATG GGTTTATTTA CGTTGATTTA GACGATGAAG GAAATGGTAC	4380
TAAAAACCGT TATGAAAAAA AATCGTTTAA TTGGTACAAA CAAGTTATTG AAATAATGG	4440
ACAAAACCTA GACTAGTTAT TGTAATTTAA CAAAACAAGA TAAGCGTAGT AGGTTGTTGC	4500
TTCTGTTTCC GCTGTTTATT AATTTTAAAG CTTGGACCAA AAATCTAAAG TGATTTTGG	4560
CCCAAGCTCT TTTTGTCAA TTTTCAAAA TAGTGCACAA TACGTTATAA TAGAAACAAC	4620
TGAGAGGAGC GTTCTAATG ATTAAGCAAT TACAACATCC ATTTTGTGTT ATTATGGGTC	4680
CGAGTGGTTC TGGCAAGACC GCAATTACAA GTAAGGTTTT TCCGAAAAAC TATAAAGTTA	4740
TCTCGCATAC TACGAGGAAA AAACGATTGG ATGAAGTGGA TGGCGTGGAT TACTATTTTG	4800
AAACAAAAGC AAGTTTTCAA GCCTTGATTG AAACCAATCA GTTAGTGGAG TATGATTTTT	4860
ACCATGGAAC CTATTACGGT GTCGGTGTAG CTGCCATCGT TGAAACAACG AAAGAGCATC	4920
CAGCCTATAA TGCCTTAAC TTTCCAGGCT TTCAAGCGGT CTTTGAGCGT TTCGGCGAGT	4980
CCGTCATTCC AGTTTTTTTT GACGTTTCAA AAGAAAATAT TTATCAACGT TTAAAACAAC	5040
GAGAATCTGA TCCTAAAATT ATTGAAGAAC GTTTAAACCT TTATGATCAA GAAATTCTTA	5100
TCAAAAACCA ATTGGAACAA TATCCTAATT ATCAACGAAT AGATGCCAAT GGACCTATAA	5160
AAGAGGTGCG TGGCTTGTTA CAAGAATGCA TCAACACTA TTATTAATCA AGCAACTGGT	5220
GAGACAGAAT CTAACATCAA TTTTGTCTTG CTAGGTGCTT GATTTTTTTT GTAAAAAAT	5280
TGCATTCTGA ATAGAGAGTG CTATAATGAC ATGGTGTGAT ATGACAATGT GTCATAAAAG	5340
GGAGGAACAC AAATATGCCT ACGAATACGT TTTTTCATTT GCCAGAAGAA AAGCAACAAC	5400
GCTTATTAGA TGCTGCGCAA ATTGAATTTT CGCGCCATTC TTTACAAGAA GCGTCAATTG	5460
CAAATATTGT TAAATTAGCG GGAATTCCGC GCGGTAGCTT TTATCAATAT TTTGAGAACA	5520
AAGAAGATTT GTATTTTTAT TACTTTGCGA CATTAAGAAA AAATAGTGAA CGAGATTTAG	5580
AGAAACAAAT TATTGCGGAA AATGGTGATT TAATCGAAGC CATGGATGTC TATTTCTCCA	5640
AAATGATTGT CGAAGTGCTA ACTGGaGAGA ATGCTCTTT CTATCGGAAT TTATTTGTCA	5700
ACATGGACTA TCGTGCTTCA cGTCGGGTGA CGGGATnAAT TTTAGCCGAC TGGGGGAAGA	5760

GGGAAAAAAA ATAGGGAAAC CAGCCACTGT CCATnAAGCC TAGAnGGGAG GAAAAAGGnC 5820  
 CATGCCTGCC CCATGGCGGA 5840

## (2) INFORMATION FOR SEQ ID NO: 32:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6590 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CGATCTTTTT GACGACGTTT AATTTTACGT TTTTCTTAT TGCTGTCTTC AATTGCCCAA 60  
 TTAATTTTTT TCTGTATCC TGGTTAATT TTTTCTTTT TCTTTTTCAC TAAACCAATT 120  
 AACGTTGGGT CTAATCTTC TCGAGATTTT TCTCGTTTGT TTCGACGATT ACGATCATAC 180  
 GTCTCTACAA TCTCACCATT TTTAATTTCT TTTGGTTTGA AAGAAACACC TAACTGCTCA 240  
 ATCTGAGTAA tCGCTTCATC ATCTgCTGGT GAATATAATG TTATGGCAGT CCCATTcAAT 300  
 CCATTACGTC CTGTGCGGCC AACCCGATGG ATAAAGAAGT CCAATTCATG TGGAACTTCT 360  
 GCATTGATGA CATGTGAAAC CCCTTCAATA TCAATCCCAC GTGCGGCTAA ATCAGTAGCT 420  
 ACTACATATT GGTAGTCAAG ATTTTGAAC TACGCGATCA CACGTTTTCT TCGCGTGGT 480  
 GTAATGTCTC CATGAATTTT GGCAACTTTT AGGCCCTGAT CTTTAAAGTA ATCGGTGATT 540  
 TCATCAACTC TTTGTTTAGT ATTTGCAAAA ACAATCGCTA AATAAGGATG CCCGATAGTC 600  
 AATAATTGAT AAATTATTTT ATTGCTGTTT TTTCTTTGG TGGAAATCAA CCAATTGTCA 660  
 ATTGTTTCTG AAATAACTGC TTTTGGTTTG ATATGTTCAA TCACTGGATT TTCTAAATAT 720  
 TTTTTTAAAA ATGGTCGTAA TTTTCTGGA ATGGTTGCAG AAAAAACGAG CATTTGTAAT 780  
 TTTTCTGGCA AACGACCCGC AATTTGATCG ACTTCTGCTA AAAAGCCCAT ATCTAATGTC 840  
 ATGTCTGCTT CATCGACAAC AAAGGCAAAA GCCGTATGCA CTTTCAAGGC TTGTTCAATC 900  
 ATCATATCTA AAATACGCCC TGGTGTACCA ATCACCACAT GAGGTTGCTG ATGCTTTAAT 960  
 TTATTTAATT GACGTTGTTT ATCTGTTCCG CCAACAAAAT TTGACACTCG AATTTCTGkT 1020  
 TGACTGAAAC GAGCCAGTTG tGCGCTTCTT GATAAATTTG ATTGGCTAAT TCACGACTAG 1080  
 GCGCAGTAAT CACAATTTGT ACTTCGTCGA TAGTTGGCTT AACTTTATCC ATTAATGGAA 1140  
 GTAAAAAGT ATGCGTCTTT CCACTTCCTG TTTGCGATTG TCCAATAACA CTTTTCCCTT 1200  
 TCTTAATGAT TGGGATTAAT TTTtCTTGGA CTTCTGTTGG TTCTTCAAAC CCTTTTTCTG 1260  
 CGAGTGCTTC GTTGATAAAA GGTGAAATT GAAATTGTTT AAATGAAGGC ACTTGCCTCA 1320  
 CCTCTTCTCG TTAAATTCTT TATGCACTTT GGCATTATAA CACAAAAACA ACTTTTACTC 1380  
 ACCTGCTGAT TCCTTTGTTT TCGCAACTTC TAATGTACTG ACTGTATAAG TGTGGCTGGC 1440  
 TTTTTCATAA AACAAACGCC GTTTTTTACG AATCAAGGTT AACAAAACGT TGATCAAAAA 1500  
 TAGTAAATAT AAAATTAAAC AGAACAAATA CAGAAGCAGT GACACCATTG GCAACATCCT 1560



ATTAGAAGAA TTCAATAACT CAGTGGCATT GGCTAGATAA ATACTAATTA AACCATACAC	1620
AAAATACAAA TAACCGTAGC GAACAAACAA GGCTTTTAAA CGAATTTTTT CATGTCCTGT	1680
TTCAACCACC CGAATGCGAA CAACTTTTTT CCCTAACGTT TGACCATTTG TCAGTTTCAT	1740
CATGACGATA AAGACCGCAA ACACCATGAT AAAGAACCAC ATCCGATCTG AAAAAATCGG	1800
ATAATTGGTT GCCCAATTTT TATAAGCTGG AATGAAATTT AGTCCAAAAA CGAGCAGGCC	1860
TTGCACAAAG CCAATCACCA ACCAATCAAT TCCCCATGCA AAgAAaCGTC TTAACAGGCT	1920
AACTGTCTGC CCTTTTTCAT AAGAgGTGGC ATCAATTCT TCTCTGTTG GCAATAAAAA	1980
GGTAAAGAGC GCGGTCAACC AATAACCAAC TATGCCTCCG ACGGTATTAT TAATTAAATC	2040
ATTGACATCC GCCAAGCGAT ACGAACGTGG ATAAATAAAA TAAAGGCCAG ATAAGTGTGT	2100
TAATTCAnAA AATAATGACA ATAAAAAACT CGCAAGAACT GTTTTTAAAA ATGAACATTT	2160
GAAATAATAC CGTAAATACA CACCAAAAGG TAAACTAAT AAmACATTAA ACAGTGGTTC	2220
TAAAACGGCA TGCTGTTTTA ATGCTGGCAA ATACGTACTA GGATCTTTGA TCGAAAAAAC	2280
AGTTTCGTTT AAAATCCCTG AAATAAAATG AAAAGGCCGC AGTTCCATTG TTGGTCCTGT	2340
GTATTGAGCC ACTTCTGCTC GAGATGGCAA CGGTAAAAAT ACCAAGAAAT AGGCACATAA	2400
TAAATAAAAA ACAATGAAT ACAAATCAT TGCTCGTGTC CAAACAAATG AACCATATTT	2460
GCGGTATTCA TAGATTAATA ATGCCCCAGA AATAAAAAAT GCTAGAAACG GGAAAATAAT	2520
GATTGCCCAT TTAATTGGTT CGCTATATAC TGCCATAAAA ATCCCTCCAA GATAACAAT	2580
TTCAAGTATG AATGCCTCAA TAACACTATA CTTGTTTCTA AAAAAATGTA CCAACAATAA	2640
ATTTTATTTT GCAAATCCTC TCCTTTGCAG TATCATTAAG AAAAAGGAGG ACGTTCTCAT	2700
GTTACAGCCA TATTTTGGT TTGGAGTGCC GCTATTTTTA CTTGTGCTCT ACTTACTCTT	2760
TGCACTGATT CACCGTCAAA CAACGATTCA TTATTTACGG TTTATTTTGT TACTTATTTT	2820
AACATTTCTC ATGGTTTTTA GTTTCCAAGT ATTGCAAGAA TCATGGACCA TCAACCCTGA	2880
AACGTAAAAA GATGCCGCTT ATTCTCCGCA ATGGTTATGG ATTCCTTTAG GTATCGGGTT	2940
AATCCTCACA CTTTACAACG CTTGGCACGG CCTTCGTACA ATGATAAAAT ACAAGACCGA	3000
TAAACATTAG TAAGAGTCTA GGACGATGGT TGTGTCCTAG ACTCTTACTG TCTCATCATT	3060
CAAATGAAAT TTAATTTTTC TTAACAAACT CTGATTTTAA CTTTATTGGA CCAAAGCCAT	3120
CTACCTTACA ATCGATGTTA TGATCACCTT CCACTAAACG AATGTTTTTC ACTTTTGTGC	3180
CTTGTTTAAT TGCGCCACTG GCACCTTTAA CTTTCAAATC TTTGATTACA GTAACACTGT	3240
CACCATCCGC TAAAAGATTG CCATTTGAAT CCTTAACTAC TAAGCCTTCT TCTGCAACTT	3300
CTTCTGTTGG TGACCATTCA TGACCGCATT CTGGACAAAT AAATAATCCG CGATCTTCAT	3360
AGGCATATGC CGACCCACAT TCAGGGCAGT TTGGTAATTG CTCTGTCATT AACGTTCTCC	3420
TTTGTAaaaa ATAATCTTTT TTCATTTTAC CTTATTTTCT AGTGAATGAA AAGTAGTCCT	3480
AATGATTTAT CAACGTCCGA TAACTACTTT CTAATTACTT ATTTATTTTT GAGCCCACTC	3540

TTGGCAAGCT GCTTGGATTT CTTGGTAAAT CACTTCAATT TCAGCCGCAT CACGTGCATT	3600
AGCGCCACTT GCCAAAGGAT GGCCGCCGCC ATGATGTTTT TTAGCAATTG TATTAATGAT	3660
TGGTCCTTTG GAACGTAACC GAACGCGATA ATAACCTTCC GGCTGTTCAA CAAAGATTGC	3720
CCATGAGAGA ACATCTTCAA TTTTTCCTGG CATTGAAACA ATGGCTGCCG TTTCAGAATC	3780
GACAATATCA TAAGTGTTA AAATACTTTG TGGTAAAACG ACACGAGCAG CGCCATTTGC	3840
GTGATTTCT AAATTTTGAT AAATATAACC CGCTAATTTA GCAACCTTTA AAGGCATTTG	3900
GTCTAACTCA CGATTTAAGT CTGCGGCATT AAAGTTAAAC GTACGCAATT GCGCACGACA	3960
GCGAGtGTAT GCGCCGAaGT GGaAGGATAA AGAAAACGTC CTGTATCCCC CACAATCCCA	4020
GCATACAATA AACGAGCTGC ATTATCTGTC ATCGTTAATT CTGCTAAATG CTGTgCCAAA	4080
AATCAACAAT AATTTCACTA CAACTACTTG CATTTGTATT CACCCAAACT AAATCTCCAT	4140
AAGGTTTCATC ATTCGGATGA TGATCGATTT TAATTAATTG ATCGCCTAAA GAAAAACGCG	4200
CATCACTAAT TCGTGGAGCA TTTGCAGTAT CTGTTACAAT GACTAATGCA CCTCGATAGA	4260
CATCGTCAGT AATAACGTCC ATTTAGCTA AAAACTCTAA GCCTTCCACT GTCCGCCTA	4320
CTTGATAaAT ATTTTTTyCA GGAAAGCTCG CTCTTAAGAG TTCTGCTAAC CCAACCTGTG	4380
AGCCAATTGC ATCTGGATCC GGCCTTTGAT GTCTATGAAT AATAATTGTT TCATATTGTT	4440
TAATTGCTGC CATAATTTCT TTTACTACGT CCATTCTTTG TAGCCTCCTA CGTTCTCTCC	4500
ATCACTTGGC AAACAACAAT TGCTTTTGCC ACAATGACAT TTTCTAAATA AACTTCAATA	4560
TCTAATTTtTG CAGAGCGACG GCCAATTTCT AAAATCCGTG GACGAATGTC CAATTCACTT	4620
TCCAACCTGGA TTAGCCGCAA ATAATGTAGA TTCACTTGTT CAATTAGTAC ATTTCTTTTT	4680
TGATTCATCA ACATTGTTTG TTGTGTCACG TTAGAAATAA TTTCACTTAG GACACCGAAC	4740
GAGATGGTAC CAACACTGTT AACCATTGTT GGGGCAACCG AAAAACGAAA ACGGGTTGTT	4800
TCTTTGCCAT TTTCTTCAAC AGGCATCACT TCCCCGAAA TTTGATCTGA AATTGTATCT	4860
GCAATTTGTG GTTGTGTTG CACCAATTGC ATAGCTTTCA TCACATCTTG GCGAGAAACA	4920
ATCCCAACAA GCGCTAAATC ATCTTCCACC ACCGGCATGA CTTCTAGACC ATCCCAATC	4980
ATTTGATGAC TGACACTCGC TACACTCATC ATCTTTTTCA CAACATTGGG ATCCTTTGTC	5040
ATGACTTTAT CCACGGTTAA ATGTTCACTT TTACCAATGA CATCTTTGGC TGTGACAATC	5100
CCCACTAAAC GTAATTTCTT ATTTACTACT GGGAAACGAG AATGATGGGT ACTTTCAGAT	5160
AATGCCTGAT AATCTGCAAT CGTCTGAGTT GTCAATAAAT AATTGGtTTT TTCAAGCGCT	5220
GTATAAATAT CACTGACGAG TAAAATGTCT TTTTAAATCA GCTGATCACT TAAAGCACGG	5280
TTAATCATAG TAGCCACAGT AAATGTATCA TATGTGGTTC TCAAGACAGG CATACCCAGA	5340
CGATCTGCTA AcTGCgCAAT CTCTTCAGAT GTTTCAAATC CACCTGTAAT TAAAACGGCT	5400
GCCCCATTTT CTAAAGCAAG CTTTTGCACA CcTTGGCGGT TACCAACAAT CATTAAAGGAT	5460
CCGGGAGTAA TATATCGCGT CATCGCGGTT TCTGTCATCG CACCAATTAC AAATTTATTA	5520

AGGACTTTGT CTAGACCAGA GcTACCGCCT AACACATCCC CTTCAATAAT CCGGACCACT 5580  
 TCACCAAATG TTAATTTTTC AATATGTTTT TTTAACTTCC GCTCAATTCTG TATCGTACCA 5640  
 ACACGTTGAA TGGTGGAAAC AAGCCCAATA TTTTCTGCAT CTTTGATGGC CCGATACGCT 5700  
 GTTCCTTCAC TAACACTTAA ACTTTTGGCA ATGCTTCGAA CAGAGATTCTG ATCGCCAATT 5760  
 GGTAAGTCTT CAATATGTTT TAAAATTTGA TCATGTTTCG TCGCCATAAT TTCCTCCTAA 5820  
 ACTAGGATCG TCTCTCCTGG GTTTAATACT TTCCC GCCAA CTACTTGCTC TACGAATATT 5880  
 GCTGGATTTT GTTGATAAC AGGAAAGGTA TTGTAATGAA TTGGAATGGT TATTTCTGGGA 5940  
 TTTAAGTAAG CGATTGCTTC TAGCGCATCC TCTGGTCCCA TCGTATAATT ATCGCCAATT 6000  
 GGTAAGAACG CTACATCAAT GGATTTGTCT TTTGCAAAAA GCGCATATC TGAAAACAAC 6060  
 GCCGTATCAC CAGCGTGATA AATTTTTTtA TCTTCTGCTT CTAAAATAAT CCCACTTGCT 6120  
 TCGCCCATGT AGGTCATTAC GCCGTCTATT TCATAACCGG AACTATGCAA AGCTGGCACA 6180  
 AATTTAACCG AACCAAAAGG AAAGACATAA CGCCCACCTA GATTCATTCC ATGCGCTTTT 6240  
 ACACCTTGGC ACTGCGCATA CATAGCCACT TCAGCTATCG CAATAATCGT TGCCTTGTTG 6300  
 GCTTGAGCAA TTGCTAGCAT ATCGCCAATA TGATCGCTGT GTCCGTGGGT GATTAAGATG 6360  
 TAATCAGGGT GTAAATCTGC AAGTGAAACG TcCGCTAAAG GGTTGCCATT AATAAAAGGA 6420  
 TCAAATAAAA GGTTCTGTCC GTCGTTTAGA CGAATTTCAA TACATGAATG TCCATGTCCT 6480  
 ATCAGTTCCA TTTGTATTCC TCCTACGGTA TTTTCTACAC AATCTTATGA TAATTAATCG 6540  
 kCAAATTCAT yCTaAAAGTA GkaCaGTTaG TACAGTTACC TTCCTTAATG 6590

## (2) INFORMATION FOR SEQ ID NO: 33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7657 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

ATAAAGAGGT AAATGAAGGA GAAATGCGAT GAAAAAAGA AAGGCAAGAT ATTTGAAGTT 60  
 GGTATGTATT GGAATTTTAG GGATTTATGC GGGCGCCAGT TACTACAATT CGGAACCATT 120  
 AATCGCTCAA GAAGAGACCG AAACGAGGGA CGAAAGTCGA CAGCTATCGC TGGAAGAAAA 180  
 AAGAGCGGAC CAAGCCAATC AAGACCAAGG AACCCTTACT TCTTCAACGC TGAACGAAAC 240  
 AATAATAAGT TCGTCTGAAC AACAAGAAAG TTCAACACAA AATGACAGCA GTGACCATCA 300  
 GGAAAATAAT GTGGAACCAT TAGAATCTGA AAATAGTTTA CCGCCACAAG AAAAGACAAT 360  
 TGCTAGTACT AAACCAACCG AATCGTTACA AAATGAGCAA CCAGTATCTG AAGAGCGTAA 420  
 AGCGCTCACT GAGAAGTCAC CTCATAATCA GGAACCACAA AATTCTACAT TAACTGAAAA 480  
 GACCAAAGAG CCGTTGGTAC ATTCAGAGGT AGACAATAGT CAAGAAAAGG AAAAGCAAGC 540

AGGAATTCAA CAACAAGCCG AGGCAAATTT TGTTGTCCAA GAGAATGCTT CTACTTCTGA	600
GTTTATTAGA AAAATCGGTG AAGAAGCACG AGTTATTGGT CAGCGACATG ATCTATACGC	660
TTCTGTAATG ATAGCTCAAG CTATTTTAGA GAGTGGCTCG GGCAATAGTG CGTTAGCGGC	720
GGCGCCAAAT TATAATTTGT TTGGCATTAA AGGAGCTTAC CATGGTCAAA GTGTCAGCTT	780
TCCCACACAA GAAGATGATG GTAAAGGGAA GATGACCACC ATTCATGCCG ATTTTCGTCA	840
GTACCCATCC TATAAAGAGT CATTAAACAGA CTATAGTAAA TTAATCATCA ATGGGCTGGC	900
GGGTAATCCA ACGTTTTATC ACGGCGTTTT GAAAGCAAAC ACGACGAATT ATCAGCAGGC	960
AACAAAATTT TTAACAGGTC GTTATGCAAC GGATACATAT TACGATAAAA AATTAAACGC	1020
GTTGATTGAA ACCTATCAAT TAACTGAATA CGATCAAGAA AAGAAGAAAC CAGTAGTGAC	1080
TAATTTGGCA GAAGAAAAGA AACCTTCGTT TGATAAGGAA GCAGTCAAAG AACAGCTAAA	1140
AAAAGAAGCA GTCGTCTATG AAGTAAGCAA AGGAGATTCT CTCGCCACCA TTAGTCAGAC	1200
GTTTGGTGTG TCAGCAACAG CCATCTTAAA ACAAATTTCA AAGACTCAAG AAATGTTTTA	1260
TATTGGCCAA AAGATCACGA TTCCTCAACA TACAGCGGCG ACTTCTCTAG AACCGAAAGA	1320
ACAAGCGTTA CTTCAATCAC TAATAATCAG TAAATCGGTG ACTAATGCTT TGAAGACAGC	1380
AGAGCAATCA AATGGTAAAG CGAATGACTC AAAACAGCG TCTACTGAGT ATTATGAAGT	1440
AAAAAGAGGA GAGACACTCG CAAAGATTGC TAAAAAACG GGCTATTCCT TAACCGCCCT	1500
CAAGCAGGCA AATGACTTAA CGTATTCTGT ATTAACCGAA GaCAAACGAT AGCACTTCCT	1560
AAGCTAAAGG ACTAATGCTG ATTTAACTTA TCGGTCGATG GAGTCAATAT ATTATTAGgG	1620
TAAGAGAAGA GGAGAAGTTC ATCCATGGAT GAACCTCTCC TCTTCTCTTA CTATGTTTTT	1680
CATAAATACC TTAATAAATA GTAAAATTAG CAAGAGAATT GTGTCTTTAC ACAAATAGA	1740
TAGGTTTTCT TGCCGAAGTT TGATAAAATG ATTAGTTGAA AGGATGCGTG AAAATTGAAA	1800
GAATTATTTT GTAAATACCG AGAAGTACTG GCTTATTTAT TTTTGGTGG TGCAACGACG	1860
GTCGTCAACT TAGTTGTGTT TTTTGTGTTT CAAAACGTGT TAGGGCTGGA TTACAAGATA	1920
AGTAACGCAA TTAGTTGGTT TTTATCTGTT CTATTTGCCT TTTTCACGAA TAAATATTGG	1980
GTATTCGCAA GTAAGCATGA AAGCATAGCT GGTTTTTTTA AAGAAATGGG CTTATTTTAT	2040
TGGTACCGAA TTTTATCTTT TGTGGCTGAT ATGGGTTTAA TGATTCTCCT AATTGATGGC	2100
ATTCATTTTT CATCGTTTTG GCGGAAAATG ATTACACAAG TTGTAGTAGT GATTTTAAAT	2160
TATTTCTTCA GTAAATTTTn CATTTTTaAA GAAAAAGAGG tCTAAGCAGA TTTtAGGCTA	2220
AAAAAATTT TTTTGTtATA ATATAAACAA GTTTTTAATA CAAAGAATAG AGGTAGAAAa	2280
CGATGGACTA TACATATGCA GATGATAGTT TGACATTGCA TACAGATATG TATCAAATCA	2340
ATATGATGCA AACATACTGG GAATTAGGGA GAGCAGATTT ACATGCTGTT TTTGAATGTT	2400
ACTTTAGAGA GATGCCCTTT AATCATGGct ACGCAATTTT TGCAGGTTTA GAACGTTTAG	2460
TCaATTATCT CGAAAATTTA ACGTTCACAG AATCTGATAT TGCATATCTC CGAGAAGTTG	2520

AAGAATATCC AGAAGATTTC TTGACGTATT TAGCAAATTT TGAATTTAAA TGTACTGTAC	2580
GCTCTGCTTT AGAAGGGGAT TTGGTTTTTA ATAACGAACC GTTAATTCAA ATTGAAGGTC	2640
CTTTGGCACA ATGTCAATTG GTAGAACTG CCTTGCTAAA TATGGTGAAT TTCCAAACAT	2700
TAATTGCGAC AAAAGCGGCT CGAATTAAAT CCGTGATAGG CGATGATCCT TTATTGGAAT	2760
TTGGAACACG CCGGGCACAA GAATTAGATG CTGCCATTG GGGCACAGA GCCGCCTATA	2820
TTGGTGGGGC GGACGCGACA AGTAATGTTC GTGCTGGTAA AATTTTTGGT ATTCCGGTTA	2880
GTGGGACTCA CGCCCATTC A TTGGTTCAGT CGTATGGAAA TGATTATGAA GCATTCATGG	2940
CGTATGCCAA AACACATCGG GATTGTGTTT TCCTTGTGTA TACGTATGAT ACTTTGAAAG	3000
CGGGAGTGCC AAGCGCAATT CGAGTAGCTC GAGAAATGGG GGATAAAATT AATTTCTAG	3060
GTGTGCGGAT TGATAGTGGC GATATGGCTT ATATTTCTAA ACGAGTTCGT GAACAACTGG	3120
ACGAAGCAGG CTTTACTGAA GCAAAAATTT ATGCATCGAA TGATTTAGAT GAAAAACGA	3180
TTTTGAATCT AAAAATGCAA AAATCAAAAA TTGATGTATG GGGCGTGGGT ACCAAGCTGA	3240
TTACCGCCTA TGATCAACCG GCTTTAGGAG CGGTGTTTAA ATTAGTTTCA ATTGAAGGTG	3300
AAGATGGTCA AATGAAAGAT ACCATTAAAC TATCTAGCAA TGCTGAAAAA GTGACAACAC	3360
CAGGAAAAAA ACAAGTCTGG CGCATTACAC GTAAATCTGA TAAAAAGTCA GAAGGGGACT	3420
ATGTGACACT TTGGAACGAA GATCCTCGCC AAGAAGAAGA AATTTATATG TTCCATCCCG	3480
TTCATACCTT TATTAATAAG TATGTCCGTG ATTTTGAAGC ACGGCCAGTA CTCCAAGATA	3540
TTTTCGTTGA AGGAAAACGT GTGTACGAAT TACCAACTTT AGATGAAATT AAACAATATG	3600
CAAAAGAAAA TTTAGATTCA CTTTGGGAAG AATACAAGCG GGACTTGAAT CCGCAGAAAT	3660
ATCCaGTGGA CCTTTCAACA GATTGTTGGA ACCACAAAAT GAATTTATTG GAAAAAGTGC	3720
GTAAAGACGT TAAACATTTA ACGGAAACGG TGAATAAGGA GGCCTAGAAA TGACAACACT	3780
TCAAGAAAAA ATCATTCAAG AGTTAGGTGT GTTACCGACA ATTGATCCAA AAGAAGAAGT	3840
TCGTAAAAGT ATTGATTTTC TGAAAGCTTA TTAACTAAA CATCCTTTTT TAAAAACCTT	3900
TGTCTTGGGA ATTAGTGGCG GACAAGATTC AACATTAGCT GGGCGCTTAG CACAATTAGC	3960
GATGACTGAA ATGCGCGAAG AACAGGAGA TATGTCGTAT CAATTTATTG CGATTCGCCT	4020
TCCTTACGGT GAGCAAGCAG ATGAAGCTGA TGCACAAGCC GCGCTAGCCT TTATCCAGCC	4080
AGATGTTTCT TTACGAGTGG ATATTAAACC AGCAGTGGAT GCAATGGTTG GTTCGCTAGA	4140
AAACGCTGGC GTTCAAATTA GCGATTTTAA TAAAGGGAAC ATGAAAGCAC GTCAACGTAT	4200
GATTACACAA TATGCAGTAG CTGGCGAAAA TGCTGGCGCA GTAATTGGCA CAGACCACGC	4260
AGCCGAAAAT GTGACAGCTT TCTTCACCAA ATATGGCGAT GGCGGAgcGG ATATTTTACC	4320
GTTATTTCTG TTAATAAGC GCCAAGGAAA AGCTTTGCTG AAAGAAGTTG GGGCACCAGA	4380
AGCCtTGAT CTAAAAATAC CAACAGCTGA TCTAGAAGAT GACAAACCGT TAGTTGCAGA	4440
TGAaTTGCC TTAkGTGTCA CGTATGATGC CATTGATGAC TATTTGGAAG GCAAAAAAGT	4500

CTCGGAAACA GATCAACAAA CAATTGAAAA CTGGTATAAA AAAGGACAAC ACAAACGGCA	4560
TTTACCCATT ACAATTTTTG ATGATTTTTG GAAATAAAAA CACAGACCAT CGGCGAATGA	4620
ACGATGGTCT GTGTTTTTTT ACATAACTAC TTTTGTAGAA GTAAACGTTG CTACAGCAAC	4680
TATCCCATGA TAATGAAAGT CGTTTTCATT TAGGCGTTTT CACTTCTTT TTCAAAACAG	4740
AAAAGGTACG CTCAATAATG TAAAGTGAAT GATGGGAGAG AAAAAGAGAT GAAGCATGTA	4800
ACAAAATtGG GGATTACAAAT tATAACAGGA GTTTTGGCAT TATTATTTGA ATTTATTTTA	4860
CATCAGCCGA ATTGGGCGTA TGGCATTATT TTAATAACAG GTTCTGTAAT GGCCTTAATG	4920
ATGTTCTGGG AAATGATTCA AACCTTACGT GAAGGAAAAT ATGGTGTCGA TATTTTAGCG	4980
ATTACCGCTA TCGTTGCAAC CTTAGCTGTG GGAGAATACT GGGCCAGTTT GATGATTTTA	5040
ATTATGTTGA CTGGTGGTGA TTCATTAGAA GACTATGCCG CTGGAAAAGC TAACCAAGAG	5100
CTGAAGTCAT TATTGGATAA CTCGCCACAA AAAGCTCATC GCTTGAATGG CGAAAATTTA	5160
GAAGATGTTT CTGTTGAGGA AATCAATGTT GGCGATGAAT TAGTAGTAAA ACCAGGGGAA	5220
CTAGTTCCAG TTGATGGCTT GGTAAAAACC GGGACATCAA CAGTCGATGA ATCTTCATTA	5280
ACAGGAGAAT CAAAACCAAT TGAAAAAAT CCTGGGGATG AATTAATGTC GGGTTCCGTG	5340
AATGGTGACG GCTCTTTGAA AATGGTTGCT GAAAAAAGT TAGCAGACAG TCAATATCAA	5400
ACAATTGTGA ACTTAGTGAA AGAATCTGCG GCGCGTCCAG CTCATTTTGT ACGTTTAGCA	5460
GATCGCTATG CGGTACCTTT TACACTAGTT GCCTACCTAA TTGCAGGTGT TGCTTGTTTT	5520
GTTTCAAAAA GTCCGACtCG TTTTGCGGAA GTCTTAGTTG TTGtTTCGCC GTGTCCTTTA	5580
ATTCTATCTG CCCCATTGC TTTAGTGGCA GGGATGGGTC GTTCAAGTCG TCATGGGGTC	5640
GTTATTAAAT CGGGAACGAT GGTGCAAAAA TTAGCTTCTG CAAAAACGAT TGCCTTTGAT	5700
AAAACAGGCA CGATTACGCA AGGACAACCT TCTGTTGATC AAGTCCAACC AATCAATGCT	5760
GGAATAACTG CTGCTGAATT AGTGGGATTG GCAGCAAGCG TGGAACAAGA ATCAAGTCAT	5820
ATTTTAGCTA GATCAATTGT TGCTTATGCC AGAAAGCAAG ATGTCCCAT TAAAAATATT	5880
ACAGATCTAG CGGAAGTTTC TGGTGCTGGC GTGAAGGCAT TTGTGGATGG TGCTGAGATA	5940
CGGGTAGGTA AAAAGAATTT TGTGACACAA GAGTCTCAAG AACTGAAAA AATTGATAAA	6000
ACGACTATTC ATATTTACAG TAATGGCACA TATTTAGGCC GAATTACTTT TACAGACACT	6060
GTACGCCCAG AAGCAAAAGA GACTATGGAA AAATTACACC AATTACATCT TCAACGAATT	6120
TTAATGCTGA CGGGGGATCA AGAATCCGTT GCAGAAACGA TTGCTGCAGA AGTAGGAATT	6180
ACCGAAGTAC ATGGGGAATG TTTACCACAA GATAAATTAA CTATTCTAAA AGAATTGCCT	6240
AAAGAAAATC ATCCAGTCAT CATGGTAGGA GATGGTGTA ATGATGCACC TTCGCTTGCT	6300
GCTGCAGACG TAGGTATTGC TATGGGTGCT CATGGAGCTA CTGCGGCTAG TGAAACTGCT	6360
GACGTTGTTA TTTTAAAGA TGACTTAAGT AAAGTCAGCC AAGCGGTCGA AATTGCCCAA	6420
GATACCATGA AAATTGCCAA ACAATCTGTA TTAATCGGAA TTTTATCTG CGTTTTACTA	6480

ATGTTAATTG CTAGTACCGG GATCATTCCG GCGCTAATCG GGGCTATGCT ACAAGAAGTC	6540
GTGGACACTG TGTCAATCTT ATCTGCTTTG CGTGCTCGTC GAATTGGCCA GTAAGGTGGA	6600
ACAATTAATT GCCTGAATGT CCAGCTTAGT GGTAAAwTAA AGGTACGTTT TTAATATTAA	6660
TAAGTTaCAG GGGAGAGATT GTGCAATGAA AGGAAACCAA ATAGATGGGC AACCAGTCAC	6720
GGAAGAACAA CTATTAAAAG AAGGCTACCG TAAATATACT GGGGAAGGAA TTGATATCTT	6780
TTATAATAAA GATATTTJTG AACATATTGG TAATTGTGTT CGTGGAATC CAGAAGTCTT	6840
TGAAGTTGGT CGGAAACCAT GGATTATTCC AGATAACGGA AGTGTTGAAA ATGACATGAT	6900
TGTGGTAGAT TCATGTCCAA GTGGTGCTCT AAAATATATT CGGAAAGCAG GGAACGAATA	6960
TGCAAATTAA AGAAGAAAAC AATCGTTTTG CTCTATATAA CGATGAAAAC CAAGAAATTG	7020
GTGAAATGAC TTGGTCAGAT GCTGGCGAAA GCATGATGAT TATTGATCAT ACGTTTGTG	7080
ATCCAACCTA TCGTGGACAA AAATTAGCTG AAAAACTTGT TTAGCCGGT GTTGAAAAAG	7140
CGAGAAAAGA GCATAAAAAA ATTATTCCGC TATGTCCATT TGCCAAAAAA GAATTTGATA	7200
CTAAACCTGA ATACGGAGAC GTATTGAGAA AATAAGAAG AACTacAAC AGCCGCAAGT	7260
AGAAATTTGA TTCCTACTTG CGGCTGTTGC TATTAGAGAC AAATAAGGAA AGACATCAA	7320
TTTCACTGGT GTTTACAGAT AAAATATGCT ACATTAAAGT AGATGAATTT CGCATATTTG	7380
TATAAAAAGG AGAATTAACG CATGTATAAT CTAATCTTAA CGtTGGTCAT CATTTTATCT	7440
GTTTTAATCG TCATTACTGT TATGAnGCAA CCAAGCAAAC AAAACAGTGC AGCAAGTGCC	7500
TTCACTGGCG GTGCAGATAA ATTATTTGGG AAGCAAAAAG CGCGCGGTTT TGAAGCTGTT	7560
ATGCAACGCT CAACAGCAGT TCTTGGCGCA ATTTGGATGA TTTTATTATT TGTTTTGGCG	7620
TTCTTATCTT CAAAATAAGT ACGTAATCTT TTAAAAT	7657

## (2) INFORMATION FOR SEQ ID NO: 34:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7434 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

TGTCAATGAA ATTTTCTTGA CAGCCCTTTC TTATTTCTTG TTTTCTTTC GTTTTTTTTA	60
TAATTGTTTT TCTGTTGAAG ATAGCTTAGC GCTTATTTAA TTAAATAACA AAATTTTGT	120
GTGAAATAAC ATAGATAATT TAAAACTAAG ACATAATTTT ACTATTTGCG CCCAACTCTT	180
TATTTACTGA ATACGTGTTT TCTGAACATT TCGTACTACT GTCTTCCTTG TAAAAAAGC	240
TGAGACAAAC TGACGAATCG GTTTGTCCCA GCTTTGCGTC CGAGTACATA TTGCTTAAAA	300
TCAAATGATT GAATTAATTA GACTTAGATA GCTGTTGtTG CTCCGCGGtA AACAATACCA	360
CGACGAGGAt CaACAGTAAT CACTTCGTCA TTATTAATTA ATGAKGtKGC ATCTGCAGCA	420
CCTACGATAA CTGGAATATT TkGCGCAATC GctACaACAG CTGCATGAGA AGTTAAACCA	480

CCTTCTTCAA CAACTAATGC TGATGCTTTT TCAATTGCTG GCATGTATTC TTTATCTGTT	540
GTTTTTGTA CTAAGATAGC TCCTTCAGTC GCTTTAGCAA CTGCTTCTTC AGCAGTTGCT	600
GCCACAACGG CTTTGTCAAT AATTGCTTCT TCGCCGACAC CTTGACCTTG AACTAATTTA	660
GAACCGATCA TTTGAATTTT CATTAAAGTTT GTTGTCCTT TTTGCGCAAC TGGAACACCA	720
GCAGTGATAA TGATTAAATC ACCTTCTGAA GCATAACCTT CTTCTTGTA TACTTTTGAA	780
GCTAAATTAA ACATTtCATC TGTGCTTGAT GGkTTGTCAG CAACTGTTGC GTAAACACCC	840
CATGATAATG ACAAGCTACG TGCTTTTTGT TCTGAGAACG TAATTGCAAC GATATGTGCT	900
TnTGGACGGT ATTTAGAGAT CATACGTGCT GTATGGCTG ATTCTGTTGC CGCAACGATT	960
GTTTGGATAC CCAAGTTACG TGCTGTATGG CCAACAGATT GTCCGATTGC TTCAGTCATA	1020
TCTGTTTTGC TATATAATTT CAATGCGAAT GAATCTTGGT TAACTAATGT TTCTTCTGTA	1080
CGAAGTCAA TACGAGCCAT TGTTGAACA GCTTCTAATG GGTAATCCCC TGCTGCAGTT	1140
TCGCCTGATA ACATTACTGC ATCTGTTCCA TCGTAGATTG CGTTTGCTAC GTCATTGCT	1200
TCCGCACGTG TTGGACGTGG GTTACGTTGC ATTGAATCTA ACATTGTTGT TGCAGTGATA	1260
ACTGGTTTAC CTAAAGCATT ACATTTTTTA ATTAAAGCTT TTTGGACAAC TGGAACATCT	1320
TCTGTTGGAA TTTCAACACC CATGTCACCA CGAGCAACCA TTAAACCATC AGAAACTTTT	1380
AAGATTTCAT CGATATTATC GATACCTTCT TGGTTTTCGA TTTTAGGAAT GATTGATG	1440
TGTGTTGCGT TTTCTTCTTC TAAATTTTA GTGATTCTA AAACGTCAGA AGCACGACGA	1500
ACGAAGCTAG CTGCGATAAA GTCAATTCCT tGACCAATTC CGAAACGGAT ATCGTTAGCA	1560
TCTTTTTCAG TGATCCCTGG TAAGTTTACA GAAACGCCTG GTACGTTAAC GCCTTTTTTG	1620
TTTTTCAAGA CGCCTCGTTT TTAACAACG TTACGATTTC GTTTCATCA CGGTCGATAT	1680
CAGTTACTTC TAAATCGATT AAACCATCAT CTAAAAGAAT GTGTGAACtA CGTTTACATC	1740
ATTGATTAAT TCTGGATAAG TGATTGAGAA TTTTTCATTT GTTCCTAAAA CTTCTGTCAT	1800
AGAAATACGT ACTGAGTCAC CAATTTTCAT TGTGATTGCG CCATTTTCCA TGTCATTGTT	1860
GCGGATTTCA GGACCTTTTG TATCTAGTAA GATCGCTACG CGTTTCCCTG TAATTTTAC	1920
AGCTTCACGA ATGTTTTTAA TACGAGCACC ATGCTCTTCG TAGTCACCAT GTGAAAAGTT	1980
CAAACGGCAA ACATTCATAC CTGCATTAAT TAAGTTTACT AGCATATCAA CGGATTCACT	2040
AGCTGGTCCA ATCGTACATA CGATTTTCGt TTTTTCATT ACTAAAACGC TCCTATTCCA	2100
AATTTGTTTA TACAAAGCTA ATAAGATTAG CTAATATACA AGATTAAAAA GAAATTTAT	2160
GATTTAACTC ATAAAGAGAT AAATCTGGTT TATGTTTATT GTTTTCTAAC GTGTCAATAA	2220
TATCAGCTGC AACTACTTGG TTATCTAACA TACCGATACA TAAACCGCCT TTGCCTTCTT	2280
TTAGAAGTTC AACAGCATAT GACCCAAATT TACTTGCTAA TACACGGTCA CGAGCACTTG	2340
GCGAGCCCCC ACGAACCACG TGTCTAAAAA TTGATACACG CGTATGGAAA TCACCGTATT	2400
CAGAAAGTTT ATCAGCAAAT TCATTTCCGC CCATTACGCC TTCTGCTAAA ATGATTAAGC	2460



AATGTTTTTT CCCGCGGTCG CGGCCTTCTT GAATACGTTT GGCAACGTTT TTCATATCAA	2520
AATCATGTTT AGGAATAATG ATTTTCATCTG CTCCGCCTGC CACGCCTGAC CACAATGCGA	2580
TATCACCAGC GTTACGTCCC ATTACTTcGA TAACGAATGT ACGAACATGA GAAGTCGCTG	2640
TATCACGAAT TCGGTCAATT GATTCCAAAA CAGTGTTAAT CGCTGTATCA AAACCGATTG	2700
TGAAATCTGT TCCAGGAATA TCATTATCAA TCGTTCCAGG AATTCCAACA GCGGGGAAAC	2760
CACGTTTTGT CAAAGCCATT GCGCCATGAT AAGAACCGTC TCCACCAATA ACAACTAAGC	2820
CTTCAATACC GAATTTTTTC AATTGTTTGA TTCCTTTTAA TTGCCCTTCT TCAGTAGCAA	2880
ATTACAGGATA ACGAGCAGAA TACAAGAATG TTCCACCACG TTGAATCTTA TCGCCAACGT	2940
CAGCAACATC TAAACGACGG ATGTCGCTG CACTAAACC GGCAAATCCA TAGTTAATAC	3000
CGTAACTTC GATTCCATCA AAAATACTTT TCGGTACGAC TGCACGAATT GCAGCGTTCA	3060
TTCCAGGGGC ATCTCCCCCG CTGGTTAAAA TACCAATGCG TTCATTTCG TATTTACCT	3120
CATATAATAT ACTTTCAACT GCCAAAAGAA ACCTTTTCGC AAAAATTGAT ATACAAAAAT	3180
ATTACGTCAT TCATTTTACC ATTAAAGAAC CAAAATGTCT TGAGTTAAGA AGAAAAAAT	3240
GAAAAAGTT GAAAACAACT AGTATTTTTT ACCGTAAATC ATTTGAAAAT GACGTTTTC	3300
TCGCCTAAAA CGTAAGCTAA TTGTGCTGA CTAGCGCTTG TATGAGCGAC CCAATTTTCT	3360
TCATTTAATA CTATTTTCTT ACCTGTTTTT TCAAAATAGA CAATCACTGG AATGTGTCCA	3420
GCGTGACGAC TGAGTACCTC TTTCATTTTC TGAAAAACGT CCGTTTGTTT GACATCTTTT	3480
GTAATTCTGA TAAAACACGT TTCGTCACTA ATTGATGTTT CTACTTGTG CGCTTTTTCT	3540
ATTGTTGAA CCAATATTTG CAATCTTGA TTGTACGTAC TGC GTTCCAC TTTTCCTTCC	3600
ACATAATAAA CTGCCCCCTT CTCGACACTC TGTCTTAAC GGCGAAAAAC GGTGGGgAC	3660
AACGtTAGTG AAATACTTCC TGTTGtATCA TCCCCATCAA CAAAGGCCAT TTGTTCTCCC	3720
TTTTTGGtCC GAATCGTCCG AATGTTTTTG ACATAAAyCA ATAATTTTCgT GGGcTGATTT	3780
TCAACAaGTT CATGAACAAG TTGCGTTTGC TTGGCTAAGC GTGTTTTTTT GAACTCTTCT	3840
GTCGGATGTC CTGATAAATA GACCCCTAAA TATTGTTCTT CTTGTTCTAG TCGCTCTTCT	3900
AGCGGATAAT CGGCAATTTT TTCTTCCTTG GGCTTCAACG TTTCTAACAA ATCCATACTA	3960
CCACCGCTGA GAAGAATACT TTGGATATTA CTCTCTAAAC CTACCGCTAA TTGACGGCGA	4020
TTAGGCGCAA TTTCGTCAA TGCACCGATG GCAATCAATG GTTCGATATT TTCACTTTT	4080
AACCACTTAC GGTCAATTTC CAAAAGAAAC TCATCGAACG TTTGTACGG TCCATTTTCT	4140
CGGCGTTCAT TTAATATATT TTGAATGAAG TCTCGGCGGA TTCCTTTTAA AGAACTAAAA	4200
CCAAATAAAA TTTGCTGTTG CTGATTTAGG TAAAAGCTGT AGTGACTTTG GTTAATTGAC	4260
GGCTGCAAAA TCGTCTGCTT GTTTTTTCTA GcTTCGCCAA TATATTCTTT AATTTTTGTT	4320
GGATTATGAC GAACAGAATG CAACAATGCG ACATAAAAAG CACTCGGATA ATGAACCTTC	4380
AAGTAAGCCA TTTGAAAACC AATAAAGGAA TAGGCCAATG CATGCGAGCG GTTAAAGCCA	4440

TAATTAGCAA AGCGCTCTAT ATAATCGTAC ACTTGGTTAG CTGTTTCTTG GGGATAGCCT	4500
TGTTGTAACG CACCATTAAC AAAATGATTT CGTTCTTCAT CTAGCACATC TTTTCTTTT	4560
TTACTAATGG CCCGACGTAA AATGTCAGCT TGCCCTAAAG AGAAGCCTGC CATTTCGCA	4620
GCAACTTGCA TAATCTGCTC TTGGTAAACA ATAATTCCGT ACGTATTTTC CAAAATGGGG	4680
GCGAGACTAG GCTCTGGATA ACTAATTTGT TCCAAGCCTT TTTCCGGCG GATGaAAACA	4740
TCAATATTTT GCATCGGACC AGGACGATAC AAGGCATTGA CAGCAGCAAT ATCTTCAATG	4800
CTAGTTGGCC CTAATTTTCT TAATACATTC CGAATTCAG CAGATTCAAA CTGGAAAACA	4860
CCACTTGTTC CCCCTTTTCT AAATAAGCC AGCGTTGTTT CGTCATCTAA TGGAATCTGA	4920
TTTAAACGAA TGGTTCGATT ATAGACGCGT TTAACAGCTG TGAGGGTATC .ATCawkGATG	4980
GATAAATTC tTAAGCCCAa GaAATCCATT TTCAGAAGAC CAATTTTTTC AACATCATTC	5040
ATAGTAAATT GGGTCAATAA AATTTCATTC GATCCTGGTT GTAACGGAAC CAAATTCAAA	5100
AGATTTTCAT CaCTAATTAC CACACCTGCT GCGTGCGTAG AAACATGGCG AGGCAATCCT	5160
TCTAGCTGAA CAGCTGTTTT ATAGAGCAAT TGATTATTTG GTGAAAATT GACCAATTCT	5220
TGCATCCgTT TTGACTCTTG GTATGCTTCT TCTAACGTAA TTTTCAGCTT GTTGGGGACA	5280
GCTGCCGACC AACGATTCGC TTTACTCTGG GACAACCCGA AAACACGTGC AAcGTCCCGT	5340
AAAACCATTT TCGCAGCCAT CGTACCAAAC GTGGCaATTy sCGCCATGTG ATAATGACCA	5400
TATTTTTGAC TGACATAAGC TAAAACTTCC TCCCGCCGAT TATCAGGAAT ATCTAAATCA	5460
ATATCGGGCA TCGTGTAGCG TTCCGGATTC AAAAAACGTT CAAAGAGCAA ATCATATTTG	5520
ATTGGATCTA CATCTGTAAT TGATAGCACA TAGGACACTA AAGAGCCTGC AGCTGAACCT	5580
CGGCCAGCGC CTGTAACAAT TTGATTGTTA TGACAGAAGT CCATCACATC CCAAACGATT	5640
AAAAAGTAAT CATCAAAGCC CATATTGTGA ATGATGGTTA ACTCTFTTTC TAATCGTTTT	5700
TGGTAAACTT CATTTAGATG AGGCACTCTT TCTGGTAATT TTTCTAAACA CAATGTGCGT	5760
AAATAAGAAC CAGCCGTTTG ATTATCAGGC ACAGGAAAAT GCGGTAATAA TTTTGTATGT	5820
AAAGGAATGT CAACTCGGCA ATTTTCAGCT AGCTGAATCG TATTCTGGAC CACTTCTGCA	5880
CCAGCTATTT GAAGCATTG TTCCATCGTC GTCTCTGCTT CACGCAAGTA ATTCAGACCT	5940
TCTATTTCCG GTGTTTGGCT TGTAATATTC AATTGATTAT TTTCTTTGAT GCTCTCAAGA	6000
ACAGTTACGG CAAAGCCTTC TTCCCTTGTT AAATGACGTA CCTCTTGCAA AGCCACGCCA	6060
GGTAATTGTT CTCTTTGGTA AAAAGAAAAG AGCGCTGGAT TACGTTGTTG ATTCATTGAA	6120
AAAGAATGTT CAACAAATAA CGACTGGGGA TCAAAATATG AAGCAAGTTG TTGGAAAGCG	6180
TGAACCGCTT GTTGTTTCATC TTCAAAAAT GTTTGGGCGA CTTCTCCTTT AGCTGACAAA	6240
GTCACGGCAA ATAAATCACT TAATAAGGAT TGATGTTCCG CTAAGTAAAA GGGACCTCT	6300
GTCGTCATTT TGCTACTAGA TAGTTGCATT AATTGCTGAT ACCCTTTTAA ATTTTtagCA	6360
AATAAAAGCA ACTCGGAATG CCCTTCTTTT TCTGCTGGG TATATTCTAA TCTTAAACCG	6420

ATAATTGGTT GAATTCCCGC TTCTTGGCAT GCTTCATAAA ATTCAACCGC CCCATGCAAG	6480
ACATTGATAT CGGTAATCGC TAATGTTTGA TAACCTAACg CTTTCGCTTG TTTTACATAG	6540
GAACGAATCC TAACCGTACT AGATAAAAGC GAATAGGAAG TAACGGTATG CAATTGTGGA	6600
AATGACATTG GACAGCTCCT TTTTGAATCC TTGCTATTTT TTCCTCTTCT CCTATTTTCC	6660
TATCGCGGAT TCTTTACTTT TAAAGAAAAT ATGCTAAACT AAGAGCTAGA AAAGAGGTGC	6720
TCATATGCGT TATCTTGTG TATTTTCTG GACCTTCGCA TTAGGCCAAG TCGTTGGATA	6780
TATCGGCGGT GCATTATCTG GTGGTTCATA CGACTTTAAA ATGACAACGA TTCTATCGTT	6840
GGCAACAGCC GTTATTATTT TATTAATTGC aCACTTTGCA GTACCGAAAG AAACCAAACC	6900
TGTTAAATAA GAAAAACCA GTTACTAAAG TTATTGTAAC TGGTTTTTTT ATCTTTTCAA	6960
CTGTTTAGAG CTTTTATTT AAAACTTTTC GGAAATTAAA GAAGGACAGC CCAATGTTTC	7020
CTAAAACGAA ACAAGCGGT ACCAGAAAA CAGAAGAATA GCCAAAACCA TGGGCAACTG	7080
TCGACCCTAC CATCGGTCCT AAAACTTGAC CGAAATTACT ACACATCTGA TTGTAACAT	7140
AAACTCTACT TACGCCCTCT GTTGGCGTTA TTTTACTAAT TAACGTATTA ATTGATGGCA	7200
TTAAAGCACC TGTTGAAAAA CCTAAAAGGA ATCGTAAAC ACCTAACTGA AAAGGTGTTT	7260
GCACAAACGC CATGGGAATG TAACAAATAA AAGAAAGAAT TAATCCGCCT AATAAAACTT	7320
TTTGGTCCC AATGCGGTCC CCAATTTTTC CCAAAGTTGG AGAAGAAATA ATTnnTGAA	7380
CGCCGGCAAT CGAAACAATC AAACCACTAA CAAATAATAC ATTCTCTGTA TCGC	7434

## (2) INFORMATION FOR SEQ ID NO: 35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2732 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

TCGGAACAAG CTTTAAAGT AAATTGAGAT TTTATATAGT GAAGAATTG CTATTGCTCA	60
ATGGGCTGTA GAATATTTAA CACAAACCTT AGAGATTCCA TTTAGTTATG ATGAAGCGGG	120
GTATATTGCG ATTCATATCC ATAGTGCTCG CAGCGGGCGT ACTGATAATA GTAAAAGTAT	180
CCGTGAAGTT ACAATCGTTT CTGAAATTAT TCATTTAATC GAGCAGGAAT TGGCTATTGA	240
TATTCATGAT GATAAAAATA GTCTCAGTTA TTCACGTTTG GTGAATCATT TACGTTTGT	300
TATTCATCGC TTCCAACAAA ATCAATACGC TGTTTTAGAT GAAGAAATT TGGAAGTCGT	360
AAAGAAAAAG TATGCTGAAA GCTATGAAAT CTCAAAAAA GTACAAGTAT TATTAATGAG	420
AAATTTTCAT TATCAAGTAC CAAATGAAGA ACTTGGTTAT TTATCGATTC ACATTGAACG	480
ATTAAGAATG ACCAAATAAT AAAAATTAAA TAGTTTATGT GAGGAGTTAC GTATAATGAA	540
AGCGTATATG CAAAGAATGG GACGTTTCGT AATGTTACCA GTTGCTGTAT TACCGGCGGC	600

TTCTTTATTA GTCGGTATTG CTAAC TGGAT CGTGGGAACA ATCGGTGCTA GTCCAGCCAC	660
AACTTTTCTA ATGAATGGTG GTTTGGCTAT CTAAACAAC TTAGCATTAT TATTTGCGGT	720
TGGTTTAGCG TTAGGAATGT CAAAAGATAA AGATGGATCG GCAGCATTAG CTGGTTTAGT	780
CGCTTATTTA GTGCCAAAA CTGTTT TAGC GCCTGCTTCT ATCCAAGCCA TTAAAGGCTT	840
CAAGGATATT GCTGAAGTAA ACCCAGCCTT CAACAGCATG GACAACAATG TCTTTGTCGG	900
GATTGTCGCT GGTTAGTTG CCGCGGCAAT GTATAATCGT TTCTCTGGTG TGAAATTGCC	960
AATGGCTTTA TCATTCTTTA GTGGAAAACG TTTAGTTCCT ATTATGTCAG CTATTT CGAT	1020
GTTAGCAATC TCAGCTGTCT TATTCTTCTT CTGGCCAGTT GTTTATAACG GATTAGTTGC	1080
ATTTGGTAAA GGAATCTCTA GCTTAGGTTT CGTAGGAGCT GGCTTGATG GTTTCTTTAA	1140
CCGACTATTA ATTCCAACAG GATTACACCA TGCCTTGAAT TCTGTTTTCT GGTTTGATGT	1200
TGCTGGTATC AATGATATTG GGAAC TTTT AGCTGGTCAA CAAGCGTTAG ATACTGGTAA	1260
AGCAATTGTA GGACAAACAG GGATGTACCA AGCTGGTTTC TTCCAGTAA TGATGTTCCG	1320
TTTACCAGCA GGTGCTTTTG CTATTTATCA ATGTGCACGT CCTGAGAAGA AAAAAGTTAC	1380
AGCTTCATTA ATGTTAGCTG CTGGTTTGC GGCTTCTTT ACAGGCGTTA CTGAACCTTT	1440
AGAATTCTCA TTTATGTTG TGGCTTGGCC TTTATATGTT TTACATGCGG TCTTTACAGG	1500
TATTTCTTTA GCATTGCTG CGTTTATGCA TTGGACAGCT GGCTTGCCT TTAGTGCAGG	1560
TTTTGTAGAC TTCTTCTTAA GTTTGAAAA CCCAGTCGCG AATCATCAA TGATGTTGGT	1620
GGTTCAAGGA CTTGTCTTTG CAGCAATTTA CTA TCTCGGT TTCCGTTTTG CAATCACTAA	1680
GTTTAACTTA ATGACACCAG GACGTGAAGA AGGCGATGGC GAAGAAACGC CAGATGTTGC	1740
AGAAGGTGAC AACAAATTTG CTTCTCTAGC TCGTCGTATT TATGATGGTT TAGGCGCTGA	1800
TGCTAATGTA ACTTCTATTG ATA ACTGTAC AACTCGTTA CGCTTAACAG TGAAAGATAC	1860
AGGCAAAGTG GATCAAGCAA AAATTAAAGC AACTGGGGTT CCTGGTGTGA AAGTCATTGA	1920
TGACACCAAT ATCCAAGTGA TCGTTGGTAC GGAAGTACAA TTTGTTGCTG ACGAAATGCA	1980
ACGATTATAT AACCACCAAG CGCCAGCAAC ACCAGTCAA GAAACACCTG TTTCTCAATC	2040
AGTTGTTGAA GAAAAAGCAC CAGTTTCAAC GAAAGAACT GAATTATATT CTGTAGCGAA	2100
TGGGAAAGTT ATTCCTATTT CTGAAGTACC AGATGATGTT TTCTCAGCAA AAATGATGGG	2160
TGACGGTTTT GCTGTCGTAC CAACAGACGG CGAAGTTTCA ACACCCGTTG CTGGTAAAT	2220
TACTAGCATT TTCCAACAA AACATGCGTT AGGTATCCAA ACAGATTCAG GCATTGAAGT	2280
ATTATTACAC ATGGGTCTAG ATACTGTTGA GTTACAAGGT GGACCATTTA CATTACATGT	2340
TGAAGAAGGC CAAGTTGTAA AACAAAGCGA TAAAATTGCA ACCATTGATT TAGCGGCATT	2400
AGAACAAGCT GGCAAAAAAT CAGATTTAAT TGTTGTTTTT ACAATCAAG ATATTGTTGC	2460
TCAATACGAT TTACAAAAAG CAGGTCAAAC GACTTCAATG AATGATGTAA TCGGAAATGT	2520
TACCGTAAAA TAAAAGTAAA AAAGGGCTAG CCGTAAGGCA GCCCTTTTTT TGATGAAATT	2580

CGATAAAATT ACTTGTTATT CTATTCTAGG TGAAAACGTT CAATCTTGCG AAGTTACGGT	2640
TTGTTCACAA AAAAGTTACT TTTATTAAAT AAAACTAAAG AAGTAACAAA AACTTCGCT	2700
TTTTCAGTTA TTTCAAGTAA CATAGAACGT AT	2732

## (2) INFORMATION FOR SEQ ID NO: 36:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1150 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

CCAATGGCTG GATGGAACAG AACAGACGT GGTATCACGA TTACATCCGT GCGACAACCTG	60
CACAGTGGAA GGCTACCGTG TAATATCATC GATACTCCAG GACACGTGAC TTCACAACTG	120
AGTACACGTT CTCTACGCGT ATTAGATGGT GCGGTAACAG TTCTTGACTC TCAATCAGGT	180
GTAGAaCCTC AAACAGAAAC TGTTTGGcGT CAAGCGACTG AATATAAAGT TCCTCGTATT	240
GTATTCTGTA ACAAATGGA TAAATCGGT GCGGACTTCT TCTATTCAGT AGAATCATT	300
CATGATCGTT TACAAGCAAA CGCTCATCCA ATTCAAATTC CAATCGGAGC TGAAGAAGAC	360
TTCAGTGGAA TCATTGACTT GATTAAAATG AAAGCTGAAA TCTACACAAA TGATTTGGGA	420
ACAGATATTC AAGAACTGA TATCCTGAA GATTATTAG AAAAAGCGCA AGAATGGCGC	480
GAAAAATTAG TTGAAGCTGT TGCTGAACT GACGAAGACT TAATGATGAA ATATCTTGAA	540
GGTGAAGAAA TCACTGAAGA AGAATTAGTC GCTGGTATCC GTCAAGCAAC AATCAACGTT	600
GAGTTCTTCC CTGTATTAGC TGGTTCAGCA TTCAAGAATA AAGGGGTTC ATTAATGTTG	660
GACGCTGTTT TTGATTACTT ACCATCTCCA TTAGATATCG ATGCTATCAA AGGGATCGAT	720
ACTAAACAG ACGAAGAAAC AACTCGTCCA GCTGATGACG AAGCACCTT CGCTTCATTA	780
GCATTTAAAG TTATGACTGA CCCATTCGTA GGTGTTTTAA CTTTCTTCCG TGTTTATTCT	840
GGTGTCTTGA AAAGTGGTTC ATACGTATTG AACGCTTCAA AAGGCAAAA AGAACGTATC	900
GGTCGTATCC TACAAATGCA CGCCAACACT CGTCAAGAAA TTGACAAAGT GTACTCTGGT	960
GATATCGCTG CTGCTGTTGG TCTAAAAGAC ACAACAACG GTGATACATT ATGTGCATTA	1020
GATGCACCAG TTATTCTTGA ATCTATTGAA TTCCCTGACC CAGTTATCCA AGTCGCTGTT	1080
GAACCTAAAT CAAAAGCTGA CCAAGATAAA ATGGGCGTGG CTCTACAAA ACTTGCTGAA	1140
GAAGATCCAT	1150

## (2) INFORMATION FOR SEQ ID NO: 37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6729 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GAGCCTAATA AAGAAGTAAA AGCCTTCTTA GACTATATGG TCACCGATGA TGTTCAAAA	60
ACAATTGTCA AAGACTTAGG TTATTTAGCC ATCACAGACA TGCAAGTAGA ACGTGATGTA	120
AATGGAAAAG TAACTGAAAA ATAAACAAAA AACTCAGCC AATAAACAGG ATTGGCTGAG	180
TTTTTTTAGG CTTCTTTTTT TGTAAAGGC AAGGTGTTG TAAAGTGGT CCTAAGCCT	240
AAGTGAATAT CAATGGCGAT TTTACCGCCT AATAATTCTG TGTAAATTTTG GACAATAGCT	300
AGACCTAGAC CGGTGCCACC AGAATGACGG AACAGTGCTT TATCGACACG ATAGAACCGT	360
TCAAAGATCC GTTCTTGGTC TTTTGGACTA ATGCCAATGC CAGTATCTTG AACAGATAA	420
CGTAATTGCT GATTTTCTAT AGCATAAGAA ATAGTAATCT GACCATTTTC TTGGGAATAC	480
TGAATGGCAT TTTCAATTAG ATTCTTCACG ATTGGATAAA ACATTCGTA TTTGGTGAAA	540
AACTGGCTCT CTGGCAGACC ATGAACGACT GTTTTCAGCT GTTTTGCTGC CAATTGTTGC	600
TGATAAGAAC CGAGAATTC TGTAAATAAT TTTTCTAAAG TAAATTTCTG TCTGCATAA	660
GGAATTGTCG CGCTGCCCCG AGAAAGTTCT AAAATTTCTT GATTAATTG TTGCAGACGC	720
TGGGCATCTT TTTGCATGAT GTGCAAAAAT TGATTAAGGA CTCCCGGATC TTCTTTGGCA	780
CCGTCTATGA GTGTCTCGGT AAAGCCCCAA AGAGAAGTGA CAGGCGTTTT TAATTCATGA	840
GACACGTTTG ATACAAAATC TTTTTCATT TTTTCTAGTT GCGCAACGCG CSTTAAATCA	900
TAGGCAATTC CTAACACTTG ATAATCATTG GCGCTTGAT CATCCAAGTA ACGAAGAGAA	960
AGATCTAATC GTAAAGGTGT TGGTCCTGT GTACAGTGA TTTCTTGATG TAACTGTTT	1020
TTTTCAGCGA CCACTTGATG GATTAAATGA ATCAAATCGG CGCTTGAAT CACTTCGAGA	1080
TAATTGCTTT GCATATTTTT AGCAGAGAGT CCTAACATTT CTGCAATTTT AGGGTTCAAC	1140
AGACGAAGCC GTCCTTGCGT ATCAATGATA AAGACACCAA TCATCAACTC ATCTAAAGT	1200
TCGTGAAATT GCTCATCCGC AGACGTATAC GCTTTATACG TTTGACTTAC TTGTTGactC	1260
AGTAGGTAA TAGTTTCGTA AAGTTCATT CATTGAGAAG ATTGCTGGaT AATGGATCGA	1320
TTTTCATCTG GATGTTTAAG CATTGTTTC AGAATCGGTA AGACAGTCAC TAGCGGTTGA	1380
TTTTTTTGCC GTAACAATAA AAAGATAAAC AAGGTAATTA GTAGGAACAA GATGCCAAT	1440
GTAAAGATAA TATAGCGACG AAAAGATTGA ATATTTTATG AAAATAGCGC GGTGTTGCC	1500
GAAATCCGAA TAATGCCAAC GATTTCTCCT GATTTTAAAA CAGGCGTGGC AACGTACAGA	1560
AGTTCTTTAT TTAACGTTTT ACTCTTTCGT AAAGCTGAAC CATAGGTAGC TCCTGATAGA	1620
ACCGCAGCGA TTTGAGTCG TTTTGTGTC GACTCTAGCG GTTCATTTTG GCTGCTGTCA	1680
AAAAAGACGG TACCATTTTT TTGCAACAAA GTTACCCGTT CTCTGCATC ATCTAACGCA	1740
tGGGTGaCAA GGAGTTGTTG TTGGTTTGTG AGTTGCCCTT TTTCGAAAAC GGAAGGGGGC	1800
AACTGGTCTG TTAAGGGGT ATTTTTTCTT CTAAATAGG TTTCTTGCTG AGCAACCATT	1860
TCTTTTGGAA AGAAAAAATT GGTGCTGCA ATACTCCCGA CATATAATAA AAGCATGACT	1920

GCCGCAACTA AAAAATATTC AATCCGCAGT CTCTTTTTCA TCGTTTTGGC TCCTGAAAAC	1980
GGTAGCCAAA GCCACGGACA GTTACTAAGT AAACGGGCCG TTTGGGATCG GGCTCGATTT	2040
TATCACGTAA ATGACTAATG TGGACATCGA CGATGCGACT TTGTCCAGCG AAGTCATATT	2100
GCCAAATTCG TTCCAACAAG GTTCCCGAT TAATGACCCG ATCTTTGCGT TTCATGAAGT	2160
AAACGAGTAA TTCAAATTCT TTCGGTGTTA GCTCAATGGG CTGATTGCGC ACAAAGACTT	2220
CATAATTTTG TTCATCCACA CGAATCTCGC CAATCACAAG CGGA <sub>g</sub> CTTTA GGTGTGTCCT	2280
CTTGAAGCGT TTCGGTCGTG GTAGGTTTTA AACGACGGAA AATAGCTTTC ATGCGTGCTA	2340
GGACTTCTCT GGGACTAAAA GGTTTTGTTA AATAATCATC TGCCCCAATT TCTAAGCCGA	2400
TGATTTTATC TACTTGTTCA TCTTTGGCAG TAAAATCAA AATAGGCGTG TCAATTTTTT	2460
CTCGACGAAG TG <sub>c</sub> TTTGGA ATTTCTAAGC CATCCATACC AGGAAGCATT ACATCTAAAA	2520
TTATAAAATC ATATTGATTA GACAAAGCTA ATTCGAAGCC GTTTTACCA TCTTCGGAAG	2580
TAGTCACCTG ATAGCCTTCT TTTTCCAAAT TAAAGGTAA CAAAGTTAAA ATCGATGGTT	2640
CATCATCGAC GACAAGAACT TTTTTCATTT ACATAACTCC TTCAAGAAAT TTCTGAAAAT	2700
AAGTGATAT TTTACTATTG TTAGTTTAAC ATATTTTATAG AGGAAGAACG TACTTTAAAG	2760
TAAGCGAGAA TGCTACTTTT GAGAAAAGTC AACTCATGGT ATAATAGAGA AATGAAATGA	2820
GGAGAAGCGA ATGATTGGTA TTAGTAGTTG CTTAGGTGGG ATTTGTTGCC GTTACGATGG	2880
AAAATCTCAA GAACAAGAAG CGTTAAAACA ACTTGTA <sub>ACT</sub> AATGGAAAAG CAATAATGAT	2940
TTGTCCCGAA GTTTTAGGGG AGTTGCCTAT TCCACGTCCG CCAGCAGAAA TC <sub>a</sub> TTGGCGG	3000
GGATGGCTTT GATGTTTGGG CGAATACTGC TTATGTATAT ACGGATAACG GGGAAGATGT	3060
GACTGAGACA TTTAAAGCAG GGGCACAACG CGCGTATCAA AAATTAAAAG CCCTACAAAT	3120
TACAACCTTG ATTTTGAAAG AAAAAAGTCC CagCTGTGGC AGCCAGCTGA TTTATGATGG	3180
ACAATTTTCT GGCATTAAAA AAGCTGGAGT TGGAG <sub>t</sub> CGCA ACTGCTTATT tTATTCAACA	3240
AGGAATGACC GTTTATTCTG AAGAAACATG GCAACTTGCG GACATTTCTT TTTAAAAGGT	3300
AAAAACTTAG TAAAAAAGA TTGACTATTC CCTATGAAAT CGGT <sub>a</sub> CCTA TAAGGGATTG	3360
AaGATGGGAG CACGGCTCAT GGAA <sub>a</sub> TGG <sub>m</sub> A AAAACAAACC GCATGAATGC GTTATTTGAA	3420
TTTTATTCTGA CATTATTAAC GGAAAAGCAG ATGAATTATA TGGAGCTTTA CTATGCTGAT	3480
GACTTTTCTT TAGGAGAAAT TGCGGAAGAA TATGAGGTTA GTCGCCAAGC GGTTTATGAT	3540
AACATTAAAC GAACCGAAAA AATTTTAGAA GAATACGAAA AAAAGTTACA TCTTTATTCA	3600
AACTATGTTG TACGGAAAGA ATTTCTTGAA CAACTACAGC AGTATGTTCA AGAACTTAT	3660
CCCAAAGATA GTGAAATCAA AAAATATATT GAACAAATTC AAGAAATAGA GGAATGAAAC	3720
TATGGCTTTT GAGAGTTTAA CAAACCGCCT ACAGCAGGCA ATGAGTAAAA TCCGTCGTAA	3780
GGGAAAAGTT TCCGAAGCCG ACGTAAAGA AATGATGCCA GAAATCCGTT TGGCTTTATT	3840
AGAAGCCGAC GTTAATTTAC AAGTGGTCAA AGATTTTACA AAACGTGTCA GAGAACGGGC	3900

AGTAGGAGTC GAAGTATTAG AAAGCTTATC ACCAGCCCAA CAAATTGTAA AAATTGTTGA	3960
TGAAGAATTA ACGAAAACGT TAGGTTTCAGA AACGGTTGAA CTGAATAAAT CTCCAAAAAT	4020
CCCACAGTG ATTATGATGA CAGGGTTACA AGGGGCTGGT AAAACAACCTT TTACTGGTAA	4080
ATTAGCAAAA CACTTAATGA AACTGAAAA CGCTCGTCCG CTTTAAATCG CTGGTGACGT	4140
TTATCGTCCA GCAGCGATTG ATCAGTTGAA GGTTTTAGGT CAACAATTAG AAGTTCCCGT	4200
TTTTGATATG GGaACAGATG CTAATCCAGT GGAAATTGTT CGTCAAGGGT TAGCATTAGC	4260
AAAAGAAAAG AaAAATGATT ATGTCTTAAT TGATACGGCC GGCCGTTTAC ACATTGACGA	4320
AGCTTTAATG GACGAATTGA AACAAATTAA AGAGTTGGCT GATCCCAATG AAATTCTGTT	4380
AGTTGTTGAT GCGATGACGG GGCAAGATGC TGTCAACGTT GCAGATAGTT TTAATGAACA	4440
GCTTGGAATT ACTGGGGTTG TTATTACCAA AATGGATGGC GATACTCGTG GGGGGGCTGC	4500
GCTGTCAATT CGGGCAGTAA CGGGTGCTCC GATTAAATTT GTCGGTTCTG GTGAAAAATT	4560
AACCGATTTA GAAATTTTCC ATCCCGATCG TATGTCGAGT CGTATCCTAG GTATGGGGGA	4620
CATGTTGACG CTAATTGAAA AAGCGCAACA AGATTACGAT GAGAAAAAAG CAGAAGAACT	4680
TGCTCAAAAA ATGAAAGAAA ACAGTTTGA CTTTAACGAT TTCaTTGAGC AATTGGATCa	4740
AGTTATGGGC ATGGGACCGA TTGAAGACTT ATTAAAAATG ATCCCTGGAA TGAGTAACAT	4800
GCCTGGTATT GAAAATGTCA AAGTCGATCC AAAAGATGTG GCACGGAAAC GGGCGATGGT	4860
CTTATCAATG ACCCCTGCAG AACGTGAAAA TCCTGATCTA TTAAATCCTA GTCGCCGCCG	4920
CAGAATTGCA GCTGGTTCAG GAAATAGTGT GGTGGAAGTC AATCGTATGA TTAAACAATT	4980
TAAAGAATCC AAAAAAATGA TGCAACAAAT GTCCAAAGGG GATATGAACA TTCCTGGTAT	5040
GGATCaAATG TTAGGTGGCG GCGTTAAAGG CAAGCTAGGT AAAATGGCCA TGAATCGTAT	5100
GATGAAGAAA AACAAAAAGA AGaaAAAGAA AAAAAGATAA ACAGGCTTTA GTTAATTATT	5160
GGAAGAAACa GAAAGTGCGT CATCAGTCTC TTCGACTGAT GGCGCACTTT CTGTTTAAGC	5220
GTAGTTTTGG ATAATTGCGT TTAACGCTTC GGCAAAACCA GCGCCGTGTT GTTCATAATA	5280
TGACGTAAAT CGTTCATCGC CAACATACAT AAGGCCTAAC CCTTTGTGTG CTTCTGGCGA	5340
ATAGCTTGAC CAAGAGTACA TTAACCAATC TTTGTGTAAT TGGAAAATTT CTTGCGCAAG	5400
TTGGCTAGTC AACTTGGGTG ATTTTGTGATA CATGGCTAAC TTAGAAAATA AATCTTTTTC	5460
CACGTTCTGC ATTGCTTGCA TGTCTTTTTC CGTCAAGTTT AAATATTTTT TGTTGGCTTG	5520
TTCAAtCGTT TCGTTGCCAT ATTTTCCCG AATCTCTTTA CCATATTTTT CTtCGTTTTC	5580
TTGAACTTTT TGTTGTTTGA ATGCTTCGAA TTTTGTCTGA TCTGACATGT TTATCTCTCC	5640
TTTGtAATGG TGTAAGGTTT GTTGAACAGA GGCTAAAAGA CGATCAATTT CATTGCGCTT	5700
TTCCAATAGT TTTTGTGAT GCTCGTATAA AGCCTGTTCA ACATTAAAGT CAGGTTGCTC	5760
CAAAATTTCT TTAATTTTCAT CTAGTTTCAG ATCCAATTCA CGGTAAAAAA GAATTTGTTG	5820
CAAACGATTC ACCTCGGCTT GCCCGTAAAT TCGATAACCA GAAGAATTAA TTCTGGCTGG	5880



TTGAAGAAGA CCAATTTTCAT CATAATATCG TAACGTCCGT GCACTGACGC CAGATAGTGA	5940
AGCCATTTTT TTAATTGTAT ATTCCAATCAT TCATTCCTCC TGACAAAATC TATCATAAAC	6000
CTTTACGTAA CGAGAGGGTC AAGCAAAAAA ATAAAAAAG TGAGACAACC TGTAAAAAGT	6060
TGCCTCACTT TTTCTCGATT GATTCAATTA GGCAGTTAAG CGATTAAAGA AATCCACGTA	6120
AGCATCTACA ATTGATTGTA AGAAGCCAAC AGTGCCATCA ATAATTTTAC CATCTTCACC	6180
AATTAAGTTC GTAATACCGC CAATATAGGC TTCAGGTTGT TGTAAGTTG GAACATTTAA	6240
GAAGACTAAT GATTGTCTTA AATGATGGTT CGCACCaAAT CCGCTAATkG CaCCTGGtGa	6300
AACACTCaCC aCAAGTCctG GTTTTTayCC CCAAACGcTT TCgCCATACG GACGTGAACC	6360
GACATCCCAA TGCATTTTTT AAAACAGCTG GACTGAACG GTTATATTCT GGTGTTACAA	6420
ACATTACGCC ATCTACTTGT TTAACCTCTT CACGGAATGT TGTCCAAGCT TCGGTGGGGG	6480
TTCTTCaTC ATCCmAACT TGATTGTATA ATTCTAAATC ACCTATGTTA ATGAAAACAG	6540
GTTCGAAATC AGCTGGAAAT AAGTCGCCA TTGTTTTTGC TACCAATTG TTAAATGAAT	6600
CTTTTCTTAA ACTCCCTACA AAAATACCAA TTTTTTTGTC ATGTAAATCC CTCCATTTTT	6660
CTTTACTTA CTTTATGATA GTACTTTTTC AAAAAAGCG CAAACGTTT ATCCTTGTA	6720
ATTGTTGAA	6729

## (2) INFORMATION FOR SEQ ID NO: 38:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 707 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

AACtGTTcAC TTGtTCCATT AACCAGCAAT GTACCTAACG CTGTTTCCAC TTCATATTGA	60
TAGCTTTTCC CTAAAAATGT TCGGACAGCG ACAGTACCTG CGACAGTATT AGTAGCAGCT	120
TCTGAAACAA TCTCAATGTC CTCTGGAcGA AtCGTTGCCA CCGTCGTTGT AACATTGGT	180
TCAGGATTAG TCGTTTTcAC TATTTGATTT GTTGTGTCC GATAATGACC GGCTTCATT	240
TTTTCGACAG AGAAAAAGTT TTCAAAACCG ATAAATTCTG CAACAACTT AGTACGAGGC	300
AAACGATAAA TCGTTTCTGG TGAATCATAT TGTTCAATCA CGCCATTATT CATAATGGCC	360
ACTTTATCTG AAATAGAAAA ACATTCTTCT TGGTCATGTG TGACAAAAAC TGTCGTAATG	420
CCTAATTGTT GTTGAATCCG TTTGATTTCA ATCCGCATTG CCACTCGCAA TTTTGCATCT	480
AAATTACTCA ACGGTTcATC TAGTAGTAAT AGTTTAGGTT CGATAATCAG AGCACGTGcT	540
AGCGCAACCC GCTGTCTTTG ACCACCAGAT AATTGTTTTG GATAACGGTC GCCCAAGCTT	600
TCTAAACCAC ATACTTTTAG CATTTCCGTG ACTTTTGGT CCATGGTTG TTTGTTTTCT	660
TTGCGTAATT TCAGCCCAA AGCGACATTT TCTGCAATCG TTAAATG	707

## (2) INFORMATION FOR SEQ ID NO: 39:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4985 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

nCTCGGTACT TnCCCCATGG GTGGATGTGC CCGGTTCCAT TCnAGGTGAT TTAGCCAAAC	60
GTTCCTTATG CGTAAGCCTC TGAAGTCAGC CACCTTTACG AACGAAAAAG AGAGCGCAGC	120
AACGATTGCT TATTTAAGGG AATTGATTGA AAAAGTTGGT TTCAmCCCTA AATATTATAC	180
TGCCATTAAT TCAAGCTATG ATTTACCTTA CGATTTTTAT CGTCCAAATA AAGACCGTCA	240
TCGTACACAA ATTGAATTGA TGCAAAAAGA TGGCTCATTA GTGGAATTGG CAACGGTCAG	300
TCCCCTAGTT GCAGCACTTG CTGGTCAGAG TCAAGGAGAT GAACGTTTCT ATTCCCTAA	360
AGAAATGCTT GACCAAGGAA ATAAAAACA TTACGATTTA TTTGATGAGA CGTATCGTGA	420
GTTTTCAAGT TACATTCATA ATGGTGCCTT AGTTTTAAAA AAATAGGTTT TCTAGGAGGA	480
TTCTATGTCA ATTAAGTTAG TTGCTATTGA TATCGACGGA ACATTGCTAA ATTCACAACA	540
CAAGATTtAC CCCACGGGTC AAAGAAGCGC TCCAAAAAGC AAATGAGCAA GGTGTTTCGTA	600
TTGTTTTATG TACAGGCCGT CCTTTGCCAr gCGTGAAAGA ACAATTGGAT GAATTAGCCT	660
TATATGGTGA AAATGATTtC GTGATtACCT ACAACGGTTC GCTTGTCCTAA GCAACCAAAG	720
ATAATACGAT TATTTACGCG TATACCTTGA GTTATGAGGA TTTTtaGAAA TTGAAATGTA	780
TTCTCGTAAA GTCGGCGtTC ACTTGCATAC AATTGATGAT TCCGCTATTT AACTGCCAA	840
TCGCAATATT GGTAAATATA CGATTACGA AGCATCTTTA GTGAACATGC CTTTAAAATA	900
TCGTACGGTG GATGAAATGA CACCAGAGAT GAACATTATT AAAATGATGA TGATTGATGA	960
GCCGGAAGTT TTAGATCCTG CCATTGCAAA ATTACCATTA CATTTTACCG AAAAATATAC	1020
GACTGTtAAA AGTACGCCTT TTTACTATGA AATCATGAAT AAAAATGCTA GCAAAGGCAA	1080
TGCTCTAGCA AAATTGGCAG ACCATTTAGG CTTAAATAAA GACGAAGTGA TGGCCATTGG	1140
TGACAATGAA AATGACTTAT CCATGATTGA TTACGCTGGG ATTGGTGTTG CGATGGGCAA	1200
TGCGACAGAA AATGTTAAAA CAATTGCCGA TGTGCATACC ACTAGTAATG ACGAAGATGG	1260
TGTCGCTCAA ATTATTGAAA AAATGGTTTT AATTtAAACA TTCTTTTCAA TAACAATAAC	1320
CCTCAATCTA ATCAACATAG ATTGAGGGTT ATTCGTTATT TTATGATGCT GCCGCTAGCC	1380
ATTTTTCAAC GGGCAACGG TGGCTTCCTC TTTTtAATAA ACGGACTGCT TCTTCTGGTG	1440
CAACCCAGTG TAAAGTATTG GTCCTTTCTA AGGGTTCGCT AAGTTGCCGC CAAGTATTCTG	1500
CTACATAAAA ATATCCAGGA TTATAATAGG CAGTTTGGCG ATGGTTCGAA TAAAAATATT	1560
CATCAGCTTC CCCTAAATAA CAGCCAATTT CTACACTAAT TCCTAATTCT TCTAGCACTT	1620
CTCGGTGGAT CGCTTCTTCT TTCGTTTCGG TGCCTTCAAT TTCGCCTCCT GGTAAAAAAT	1680

AAGCTCCATT TGGCGCTTGT ACCAGAACCA TTGTGTTATT TTCTGGCTTT GAAACAATAA	1740
TGTATGCAGC ATATCTTGTT TGATAGGTCA ATGTCTCTTC TCTTTTCCA AACGTTGGCG	1800
TTTTCATTGG CGCACCTCTT CTCTTTTCAT CTATTTTACC TGAGAAATTC CTTTAGAACA	1860
AGTCAGAAAC CAAGCGTTTT CTCTACATTT CACGTTTATC TATGCTACTA TTAACGTTGG	1920
AAATATAATT AAGTTTGAAA GGGAGTTTTT AACTATGAAA ATGGCTCATA CATGTGTGCG	1980
TGTAAGAT TTAGAGGCAT CATTAGATTT CTATCAAAAA GCCTTCAATT TTGAAGAAAG	2040
TCGCCGTCGT GATTTTCCAG AAAATAAATT TACTTTGGTT TATTTAACTT TACCAGGAGA	2100
CGACTACGAA CTAGAACTAA CTTATAATTA TGATCATGAA GCATACGATT TAGGAAACGG	2160
TTATGGTCAT ATCGCCATCA GTACAGATGA TTTAGAAGGA TTACACGAAA AACATCAAGC	2220
GGCTGGCTTT ACTGTCAC TG ACTTAAAAGG CTTACCTGGA ACTGCACCAT CTTATTACTT	2280
TGTTGTCGAC CCAGATGGTT ATAAAATCGA AGTCATTCGT GAACGTTAAA AAGTTTCACT	2340
TGTTCACTTA CTGAGCTTTT TGTGTTTGAG AGCTGTCCGA GACAAAACCT GTCTCGACTT	2400
CTTCTCAAAC ACTTTTTTCA AAAAAAGTGC TACAATAGAA CGTATGAATT TATTGAGGAT	2460
GTGATGTTAT TATGACAAAA AAAATTATTG GAATCGCTGG CAATCAACTT TTGCAGGCAG	2520
CTGAAGTGTT TCACGGTAAC CAAGTGACGT ACACCCACA AGGTTTGTG AGCGCTGTTC	2580
AAGCCGAGT GCGCTTCCTC TCGTTTTGCC AATTGGCCCC AAAGAATTAG CCGCTACGTA	2640
TATACAACAA ATTGATAAAC TACTTTTAGC AGGCGGTCAA GATGTTGCCC CTGAATTTTA	2700
CGGACAAGAA CCCCATATAA AATTAGAAGA AACGAATCGT GATCGTGACG AGTTTGAGTT	2760
AGCCCTAATT GAAGAAGCCT TAAAACAAAA CAAACCTATT TTTGCGGTTT GCCGTGGAAT	2820
GCAATTAGTC AATGTTGCTT TAGGTGGCAC ATTATATCAA GATTTATCGA TGTATCCTCA	2880
ATGGTCAGTT AAGCACGGAC AACAAACCAC TCAACCAATC TTTGCGACGC ATCGTATTGA	2940
TGTTGAACCA GACAGTCAAT TATCAAACAT ATATGGTACA ACTGGTCAAG TCAACTCGTA	3000
CCATCACCAA GCACTACATA CTTTAGGAAA AGATTTACGC GTTACTGCTT GGTCTCTGA	3060
TGGCCTTGCA GAAGCAGTGG AATCAACAAA TGAACAGCAG CCTTTATTAG CTGTTCAATG	3120
GCATCCTGAA CTAATGTACG CACGTGATGC CAAATCCCAA GCACTCTTTA ACTACTTTGT	3180
TCAAAAATTA TAGTAAAAAC GGmGGGACGT CTTTCGCCCC TCCGTTTTTA CTTGCTATCA	3240
TACTTAAAAA TCAAAATTAT CTGGATCTGC ACCAAAACGA TGATCCTTAT TTAAGTATT	3300
AATTGCGGTT ATtTCTTCTG GTGTTAATTC AAAATCAAAG ACAGCAAAAT TTTCTGAAT	3360
TCTTTTTTCG TGAACAGATT TGGGAATCAC GATAATATCA TTTTGTAAGT GCCAACGAAT	3420
AATCACTTGG GCCGCTGATT TACCATATTT TTCACCAATT TTCACTAGTA CTGGTTCAGA	3480
TAATAAATTA CTTGTCTTA GAGGAGACCA AGCCTCCACG GCAATCTTAT TTTGCGCTAA	3540
ATATTTTCGT AGCGGTTCTT GTGTTAACGT AGGATGTAAT TCAATTTGAT CCACCATTGG	3600
CACAACATTT GCTTCTGTTA ATAAATCTTC CAAATGATGT TGATGAAAAT TGGAACGCC	3660

AATTGCTTTG GCACGACCAC TTGCGTAAAT TTCTTCTAAT GCTTTCCATG ATTCTTTATA	3720
TTTACCAGCA ACAGGCCAGT GAATTAAATA TAAATCAACA TAGTCTAACC CTAATTTTTT	3780
TAAGCTTTTA TCAAACGCTT TAATCGCATT TTCATAGCCT TGATCTTCAT TCCATAATTT	3840
CGTCGTCACG AAAATTTCTT CGCGGGGAAT ACCCGATTGA CGAATGCCTT CGCCCACCCC	3900
TACTTCGTTT TTATAAACGG CTGCTGTATC AATCAAGCGA TAACCCACTT CTAAAGCCCA	3960
TTTTACTGAA TTTGTTGCTT CTGAGCCATC TTCTACACGC CAAACACCTA AACCTAGGCG	4020
TGGCATTTC A TTGCCGTTTG CTAATTTTTT TGTTTTATCC AAAGTCATGT GTTCTTCAG	4080
TCCCTTCATT TTCATTACTT TTAGTATACT ACGCATTCTT CTTTCACGAA ACAATCACGA	4140
CTTGTCAGGT CATTTTCTTT ATTTTAAATA AATAACGACG CCAATAAGTC AGCACTCCGA	4200
TTAGTGCCAC AACTACAATT GTTCGA <sub>2</sub> CCG TAACGCTTCC AATCCACAGT TTAAGTGCAA	4260
GTAACCAGAT ACAAAGAGA TAAACACAGG ACCAATTAAA GGAAACTTTC AATGTACATG	4320
ATTCCGCAAT TTCGATTATC TGACTACTAA AAAACGACT CCCTAATCGT AAAGCGACTG	4380
GATTATGGAA GATGAATGTT CGTCGCTTGT CAGCCGTCAA GCGAGTAGCA CATCGCTCAT	4440
CAATAAAAT TGTTACAGGT AATTAGAAA GTAAATCATT TTTTGCCCAT TCAATTGAAA	4500
CTTTCACGCA GCACTCACAA CCTTAAACAA AATTATAAAA TATATTATTT ATTTACAAAA	4560
TTATAACATT CCTAAAAAA CATGTTAAGT CTTAAAAAA AACTTTCTAT TTTCTTCCCC	4620
TTTCGCCGAG AGATTTAGTA TACTAAAAGA AGAAATTTA GAGAAAGTG GATTTTTATG	4680
TTGATACTTG GTTTAGCCAG TACTGAAACA GTCGATTCTT CTGCTCCTGA TTTAGCAGAA	4740
GCGACCGTAA AAAAAGTGAG CGCGCTACAA CGCTTTTGGA ACAATATTAA TTGGGATGCC	4800
ATTGTCGCTA CCTTAATTGA AAAAAGTTA TCAATTTTAT TTTTAATTTT TCTATTTTTT	4860
ATTATTCAGC GAATCGGGAA GTATTTAATT GATCGCACCT ATGCAAATTA TAGTAAAAAG	4920
CAACATTTTA GTGAAAGCCG TTTAAAAACG TTACACACTT TAATCATTAA TGCATTTCAA	4980
TATAC	4985

## (2) INFORMATION FOR SEQ ID NO: 40:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1346 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CGTCTGTCCT TATCTCGCAC AGCTATTTGG AAAGCGATTA ATGAATTGAA AAAAGATGGC	60
TATCAAATTA CTAGCGTGCA AAACAAAGGG TATCGCCTTG AAAAATCAGA TGTCCTTTCT	120
GCAGAAGGCA TTCAGTTGGC ATTATGGCCA CAAACACCGC CTTTAACAAT TACCGTCCTA	180
GAAACATCTG AATCTACCAT GACCGATGCT AAGCGGGCGA TTTTAGACCA GACGCCCGAC	240
AATACTTTGA TTGTTGCTGA CATGCAAGAA ATGCCTCGGG GCAGGTTCCG CCGGCCGTTT	300

TTCGCTACGC CAGGTAAAGG GATTTACATG AGCATGGTGC TTCAACCGAA CCAAATTTT	360
GAAGAAATCG CTCAATATAC TGTGATTATG GCCGTAGCGG TCGCTCGAGC AATGGATGCT	420
CTGGCAGGCG TTCACACAGA AATTAAATGG GTCAACGATC TTTATTTAAA TGGCaAAAAA	480
GTTTGCGGGa TTTTGTCTGw AGCCATGAGC AATGTGGAAA CAGGGCAGAT TTCTAATGTT	540
ATCATCGGCA TGGGGATTAA TTTTTC AATT ACAGAAACCG AATTCCCTGA GGACATTCGG	600
CAGAAAGTGA CTTCCCTTT CCCAGATGGT GAACCGACGG CGACGCGAAA TGAATTAGTC	660
GCCAACATTT GGAATCATTT TTACCAAATT TTAGACGAAC GTTCAACGAA AGACTTTTTTA	720
GAAGAATACC GGCAACGCTC CTTTGTGATT GGCAAAGACG TCGCTTTTAC CCAAACGGG	780
CAAGACTTCC GCGGCGTTGC GACAACCATT AATGGAAATG GGAATTGAT TGTCCAGTTG	840
CCTGATGGCA CGGCCAAAAC ATTGTCTTCT GGAGAAATTA GTTGGACCA GATTGGTGAA	900
TGGCGTCGAG GATAAGTGGG GAAGTTTAGA AAAAGCAAG AAAGCACCTG TCTAAATTTA	960
AACAGATGCT TTCTTGCTCT TAAAAATCA ACTTCTTTTT GTGCTAACAT TTCGCACAAT	1020
TCTTTGGGGA AATTCGTATA GGGATTTTCA TTGATTTTCA TTTCTTCTAA AGGAGACAAC	1080
CAGTTTTTAT TCGTATTCTG AAAAGCAATA GCTACTTTTT CTTCAAGATA CTTACATACT	1140
TCTGCGCCTT TAAAATTTTCG ATAACGTCT TTTAAATGCA GAACTTTCTC TAATCTCCCA	1200
TATAAATtTT CTCTAGTCAT GTATCGATTG ACTTCCTCAA AGTGAAGTAG TAACCAGAGT	1260
TCAAAGCATT CATTTGAAAA AATTGTCTTG ATTCAGGAC GATGTCGTTG TGCACATCTn	1320
ACCGCGTTTT CAATCGATTG nGTGGT	1346

(2) INFORMATION FOR SEQ ID NO: 41:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1959 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

CCAAAATATk GaTGCAATGG TTtTGkTTyC CTGaCAATAG nTGayCGATC GTCACAAATT	60
CGTATCCTTC ATTTTTTAAC ATAGTGATCA ATCCTGATAC AGCTTTGACT GATTCATGAT	120
GGATATCGTG GATTAAGATG ATTGCGCCAT TAAAGACATT CTGTTTGACA ACATTATTGA	180
TTGCTTCAGG ATTTCTTGTT TTCCAATCTA GCGAATCAAT ATTCCATTGA ATAATTGGCA	240
TGCCAATTGT TTCAGCACTT TTTTATCAA TAGCGCCATA AGGAGGACGG ATATTGCGAG	300
GCAAGACGCC CGTAGCCTTA AAAATTGCTT TATTAGCTTT ATTCATTTCA CTTTGTAATT	360
CATCTGTCGA AAGTGTATTC AACTGTGGAT GAGAATACGA GTGACATGCT ACCTCATGAC	420
CCTCGTCGTG GACTTGTTTT GCAACATCTG GGTTTTGATC CACCATCTGT CCTAACATGA	480
AAAAAGTAGC CTTACGTTG TTTGTTTTTA GTATATTCAA CAAATCAAGT GTTGAGCTAT	540

TATTGGGTCC ATCATCAAAA GTAAGAGCGA CATACTTTTT. ATTTTCATCT AATGCAGGTA 600  
 ACGCATCTTT TATACTTTCT TCGGAAACCA GATCTGTATC AATAAAAGGC GCAATGTCTC 660  
 TGTATTTTAG AATAACTTTA GTAGTTCCAG TTGTGTTCTT TGGTAGGTCG ATGGTCAATT 720  
 CATCAGGTGT ATAAGTCATT TTACTCTCGT AAGAAATTCT ATCCATCGTT AATACAGCAT 780  
 CAATTATTTT TTTAGGCTCT TTGGCGTGAT CAAGAATGGA TTGCTGAATC ACCTGTTGAA 840  
 TACCTAAGAG ACTTCCTTCA TCCGGAATCA AATCCTTATT TGTGATTTCT TTACCAGTTT 900  
 TTTTCGAAAC ATAAATGGCT TTCTCAGAAA TACTGTCTGG CTTTTTAAAA TTTTCCTTAC 960  
 TGCGATCCCA GATAAAACTA TCTGCTTCAA TTCTATACTG ACCAATTTTA GAAGATATTG 1020  
 CTGTTGCTTT TATTCTGGCA ACTGTAAGCA CTTCTTTTTC ACGATGCTTC TTTTGTGCT 1080  
 TTTGAGCAAG CTCGTCTAAG AGCTGGTTTG CATTTTTAAA AGGAACAGTA CTTTCATAAA 1140  
 CTGGACGGTA AATAAGGCAC TTATTTTCTC CGTCATTAT AGTATCTATT TTCGTTGAAA 1200  
 TTCCACTTTG GTCTTGCTCT TCTTGATTG AAGCCATGAG ATTTTTTACA GCAGTTTCAT 1260  
 ATTTGGCTTG AGCTTCTACT TCTTGCTGTT TTAATGTATA AACAATATAG CCGCCAGTAG 1320  
 CGACAATTAC TAGAATAAGT GCGGTAAACA ATGCGAATAT AACTTGCTA TAATTAGATT 1380  
 TTTTCTTCGG TATTTTTGGT TGTTTGTTTT CGTTTCCGGT TTTGGGCTTT ACAGCTTCAT 1440  
 TCAATTGATT TGCATTTTCA CAGATATCAG GTTCTGTTT GTGACGTGA CTACGACGTG 1500  
 CATAGATATT TTTATCAGAC ATATGACTTT ATCCTTCCTT TTATCATCAT AAAGTAAAAT 1560  
 GAAACAGCGC AATTGTAATA AAAAATAATT TATAGATAAT TTTTACAAGA AAGTTTTCTC 1620  
 TATTATTCTT TTTCTTTGG CGTAGTAATT ATCGCTAATT CCTCACCATT TTAGATATTG 1680  
 TGGCTACTTA ATGTTAAACT CAATTTTGAT ATCATTAGAA TCACTAGTTT TGATTTTAGT 1740  
 TTTTGGAATT GTACCCTTCG ATAGGTTAAT CACTATGGTT GCAGTCTTTG AATCACCAAT 1800  
 AATGGTCTAG GATTAAAAAT ATAACTATT ATCCAATCTT GTTCTATTTG ATTAAATATA 1860  
 CTACTTTAAT ACCTTTAATT TGTGATCAAT TCCAACCTCA CCAAACCTCAT TTACAATTAT 1920  
 AATAATTTCT TCCCTGGGAT GGGGTGAGAC TTGCAAGGC 1959

## (2) INFORMATION FOR SEQ ID NO: 42:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3625 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

TTTTATTG GAGGGATTTA TTGTGTTAAA ACCATTAGGC GATCGGTCG TAATTAGAGT 60  
 CGCGAAAGAA GAAGAAAAAA CTGTTGGAGG AATTGTTCTT GCATCCGTTG CAAAAGAAAA 120  
 ACCACAAACA GGTGAAGTTA TCGCAGTAGG TGAAGGTCGT GTGCTTGAAA ATGGCACAAA 180  
 AGTTCGGATG GAAGTAAAAA TTGGTGACAC AGTAATGTTT GAAAAATATT CAGGAACAGA 240

AGTGAAATAC GAgGCGTAGA ATACTTAATT GTATCAGCCA AAGACATTAT TGCCACTGTT	300
GAATAATAAA CAACTTTTAC AGAaTAATAT AATAGAAKAA AAPAATTATG AGGTGAAGAA	360
ATCATGGCAA AAGAGnTTAA ATTTGCAGAA GATGCACGTG CAGCAATGCT ACGCGGAGTA	420
GATGTATTAG CAGATACAGT GAAAGTGACA TTAGGCCCTA AAGGTCGTAA CGTTGTTTAA	480
GAAAAATCAT TTGGTTCACC ATTGATTACT AACGATGGAG TAACAATTGC TAAGGAFAAT	540
GAATTGGAAG ATCATTTTGA AAACATGGGC GCAAATTAG TTTCAGAAGT TGCTTCTAAA	600
ACGAATGATA TTGCTGGTGA CGGAACAACA ACAGCGACTG TTTTGACACA AGCCATTGTT	660
CGTGAAGGCT TAAAAACGT AACTGCTGGA GCCAACCCAT TAGGTATTCTG TCGTGGGATT	720
GAATTAGCAA CAAAAACAGC AGTAGAAGAA TTACACAATA TTTCATCTGT AGTTGATTCA	780
AAAGAAGCGA TTGCACAAGT CGCTGCTGTT TCATCAGGTT CTGAAAAGT CGGCCAATTA	840
ATTGCCGATG CAATGGAAAA AGTTGGTAAC GACGGCGTAA TTACCATTGA AGAATCAAAA	900
GGGATTGAAA CAGAATTAGA TGTGGTTGAA GGAATGCAAT TCGACCGCGG TTATTTATCT	960
CAATACATGG TTAGTGACAA CGATAAAATG GAAGCTGTTT TAGAAAATCC ATATATCTTA	1020
ATTACCGACA AAAAAATCTC AAATATTCAA GATATCTTAC CTTTATTAGA ACAAATTCTA	1080
CAACAAAGCC GTCCACTATT GATTATTGCG GATGATGTTG ATGGGGAAGC TCTACCAACA	1140
TTAGTATTGA ACAAATCCG TGGTACATTT AATGTTGTCG CAGTAAAAGC GCCAGGATTT	1200
GGTGACCGCC GCAAAGCGAT GCTTGAAGAT ATTGCTAATT TAACAGGTGG TACAGTAATC	1260
ACTGACGACT TAGGGTTAGA GTTAAAAGAC ACAACTATTG AA-ACTTAGG AAATGCTAGC	1320
AAAGTAGTTG TCGACAAAGA TAACACAACA ATTGTGGAAG GTGCTGGTTC AAAAGAAGCC	1380
ATTGATGCCC GCGTTCATTT AATTAAAAAC CAAATCGGCG AAPCAACGTC TGATTTTGAT	1440
CGTGAAAAAT TACAAGAACG TTTAGCTAAA TTAGCTGCTG GGGTTGCTGT CGTTAAAGTC	1500
GGTGCTGCAA CTGAAACAGA ATTAAAAGAA TTAAATTAC GAATTGAAGA TGCATTAAAC	1560
GCAACACGTG CCGCTGTAGA AGAAGGCATG GTTTCTGGTG GTGGTACCGC ACTGGTCAAT	1620
GTAATTGGTA AAGTCGCTGC GCTAGAAGCT GAAGGCGATG TGGCAACAGG GATCAAGATT	1680
GTCGTTCTGT CATTAGAAGA ACCAATCCGT CAAATCGCTG AAAATGCTGG TTATGAAGGA	1740
TCAGTGATTG TTGACAACT AAAAAATGTT GACTTAGGTA TCGGATTCAA TGCAGCTAAC	1800
GGTGAATGGG TAAACATGGT TGAAGCCGGT ATTGTTGACC CAACAAAAGT AACTCGTTCT	1860
GCCTTACAAA ATGCAGCTTC TGTGTCAGCT TATTATTAA CACTGaAGC AGTTGTTGCA	1920
GACAAACCAG AACCAGCTGC ACCAGCTCCT ATGATGGATC CATCAATGGG CATGGGCGGT	1980
ATGATGTAAG AATAAAAAAG TATTCTAAAA GCTTACGATT GAAAAATCGT AAGCTTTTTT	2040
TTACATAAAG nATAATAAT CTAAATAATA TCTTTGCTTA TGGTGTTC TCAAATATTT	2100
ACTCAATTAA CAAAAAGAT GATAATTTAA AGGCTTCGTT TTAGCTAGAC AGCCTTGTTT	2160
TTATTGATA AAATTTAGTA ATAACAATCA TTTTTTGAA ACTAGTAGTT EGTAATTTTT	2220

ATTTATTAAT AGAATATAAA TAAAAATTAT TATATTTTGA AGGGGAATAT TAGGAAGATT 2280  
 GAGGTGGAAA AATGGAAAAA TATACAATTG GTGTCGTTCC TTTGACAATT GAAACAAAAG 2340  
 AATACTGTCA AAATTGGTTT GTTAGTAATC AAGTAGAATG TCAGATCATC CAAGGAACAA 2400  
 ATATACTTCT TCATTTGCAC CATATTGATG GATTAGTGAT TGAAGTTACT GACCACCAGT 2460  
 AAGTAAATAC TTGTTGTGAG TTATTAATGA CGATTTCGAA GCAGTCAGAT TTACCGATTT 2520  
 GGTATTTTTC TCGGACAGAA GTAATTAGTA AAGTCAATCG AATTATTTAT TTACAATTAG 2580  
 GTGCAGATGG CGTTTTTGAT TATTCTTATG ATCGTCAAGA ATGCATGTTA AGCATGAGTA 2640  
 ATTTATTGCA ACGTGTAATA AGACGTTTTT ATCCTAAATT AGCTATTGCA AACGAGGAAC 2700  
 AAACAGTTAC TAAGAATCTT TCTGAGAGAT TGTATTTAAT TCGCCCAAAT TTGAGTGTTT 2760  
 GCTTGGGAAG TGGAGAAGAA ATTCTTTTAA CAAAATTGGA ATTTTTCACA ATAGAGTATT 2820  
 TATACAAACA CGCAGGACAA ACAATTACAT ATGAAGAACT CTACAAAAAT GTCTGGAAAG 2880  
 ATACTGCGAA TGAACGAAAA TATCGTGTCG CTAATGTGAT TTTCCATTG CGCAAAAAAA 2940  
 TTGAACAAGA TGTGAACAAA CCACAATATA TCAAAACAAT TCGTTCAAAA GGCTATATGT 3000  
 TAACGGTGTA AAAGGTACGA AGGAGGAGAA AAAGATGTGG CAATTTTGA CTAAAAACAA 3060  
 TGAAGATGAG GAGTATGAGA AGATGCCAGA TTTTGAAAT GAACAATCGA TGAAAGCACA 3120  
 ACCAACGGTA ACAGAAGAAG CAATCTTTGA TGACGAGGTC GTTTATGCAA AAGAGCAATC 3180  
 CGCAAAAGTA AAACAGGTCC AGCAACAAGT GCCGACTGAA CTGTCAGAAG AGTCATTTTC 3240  
 TTTTGTGAA AAAGTAGAAG AAGAAAATCA ACTATATCAG CAATCCATTA AAAAGCTAAG 3300  
 GGAGCAATTA TTAATGCAAA CCAATGAAGT GGAAGCATT AAAAAACAAG TAGTCGAAAA 3360  
 AGATGTTCAA CTTAAACATG TTAAAGAAAC ATTAAGTGAT AAAGAAACAA CTATCACTTC 3420  
 TTTACAGAAA CAATTGTCTG AAGAAAAGAT GCAACAGAGA CAGACCAGTG aAGAGrATT 3480  
 AGACACAGCC GTTACGCTTT CTCAAAAAGA AATTGGCGAG TGTATTAGA AGCCAAACGT 3540  
 CAAGCAAAAG AAACATTAGT CAGCCAACCA CCAGTTGCAC ATTCATGAGA ATGGACACGT 3600  
 TGCACCTTAC ACGCTGAGCA GTGGn 3625

## (2) INFORMATION FOR SEQ ID NO: 43:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3758 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

CCnCCCCCCT TTTTTTCCG GTCCnAATn CCTCCGCCn TTATTTTAGG TTTCCATTG 60  
 GTAAAAATAG TATAGGrGsw AAgCatCaGT TATGTATtAt TAGTGamAGA AGGTTGCcGT 120  
 GAAGTAAAwA TAAAmAGGAG AGCCAAAGGA TGTTAaMAA TGTAAGAGTG TTCTGGCAAA 180



ACTTTTTAGA TAAACATGAA TTAGATATGT TGATGCCCGA TGTGTGGATG TTTGGTGATG	240
GCTCTTCTGA GATGGGTAAT CGTTTGGGTC AATTGGTTGT TTCTGGTCGT AAGACAGCGA	300
CATGCTCTTC TTTAGATATT TATAAGATGG AAGAAGAACA ACTCCCTAAA GCTGGTCAAT	360
ACGACATAAT TTTAGATGGT CAGTCTCAAC CGCTGGCTAT TATTCGAACA ACCAAAGTTG	420
AAATTATGCC CATGAACAAG GTCTCAGAAT CTTTTGCGCA AGCGGAAGGT GAAGGAGATT	480
TGACGTTGGA TTATTGGTAT GAAGAGCATG CGCGGTTCTT TAAAGAAGAA CTGGCCCCCT	540
ATCAGTTACA ATTTTATCCC GATATGCTGT TAGTTTGCCA AAGTTTGAAG GTAGTGGATC	600
TTTATACAGA AAAAGAATAG ACAGAGAAGC GGTGACTGAC ACGATTGAAA CGAGCAGTCG	660
CCGCTTTTTT TTTACTGATT TTTTGAAAAA GTTGCGTATG TTTATATAGA CAGGGTGATA	720
GAAAAGAGAT GTGTATTTTA AGTGTGAGG AAAAGATGGT ACAATAACCT AAGGAGAATA	780
CTAACAAGGA GTGAATGTTA TGAAAGAAGA ATTTTCATAT GAAATTTTAG AAGAAGTCGC	840
TGTGTTATCT GAGAATGCTC GTGGCTGGCG TAAAGAATTG AATTTAATTA GTTGGAATGG	900
TCGTCCACCA AAATTTGATT TGCCTGAGTG GGCCCTGAT CACGAAAAAA TGGGTAAAGG	960
CATCACCTTA ACTAACGAAG AATTTGCTGA ATTAAGCAAA ACAATTAmAT CCATGTAAAA	1020
GGAGCACCTT GTCAATGAAA AGTATGACTG GATTTGAAA AAAGACCATT CAAAATGAAA	1080
ATTATCAATT GGATATCGAA GTCAAAAGTG TCAACCAGCG GTTTTATAGAT ATTCAATTGC	1140
GTATGCCCAA GGAAGTGAAT GCCTATGAAC TGGTGATTCTG GCAAGTAATC AAACGCACGT	1200
TAAAACGTGG GCGTGTGGAA GTCTATGTAA ATCTCCAAAA AATCGGCAAC AATCAAAAAG	1260
AAGTGCAGCT GCAGTGGGAC TTAATTGATC AATTATTAAC TTCGGTAGAT CAACATTTAA	1320
AGGAAAAATA TCCGGAAGCA ACGTTTGATG CAGGGGACAC GGTCAACCAT TTATTAAAGC	1380
AAAATGATTT TGTCGAAATA GTGGAAGCAG AAATTGTTGA TCAGACATTT GAACCATTTT	1440
TTGTTCAAGC ATTTGAAGCA GCCATCGCTA GTCTTGACCA AAGCCGTGTG CAAGAAGGAA	1500
CGCAAATTAA ACAAGTCCTG CTTGACTATG TAGCAGTGTT GACCCAATCA ATTCAAGAAC	1560
TGCAAGCGTT TGTGGGCGTT TTTGAACAAG AATACCGTCA ACGTTTTGAA GCGAAATTAA	1620
ACGAATGGTT AGGTAGCCAA GTGGATGAAA CACGCTTATT AACAGAAATG GCCATTTTAT	1680
TGGAAAAAGG CGATATTCAT GAAGAATTGG ATCGTTTGA TATCCATATT GATAAGTTGC	1740
ATCAGTTGCT TGACGAAACA GAACCAAGTG GCCGAGAACT AGATTTCTTA ATTCAAGAAA	1800
TGAACCGTGA AGTCAATACG ATTGGTTCTGA AATCAAGTCC AATTGAAATA AAAAATAGTG	1860
TCGTTCAAAT GAAAACAACC TTAGAAAAGA TTCGTGAACA AATCAAAAT GTCGAATAGA	1920
GAAAATGGTT TCGTGTTTTT CCAGAAAACG CTATAATATA GAAGTAACGA ATAAATGACA	1980
GAGGTGAAAG CAATGACTAA ATTTTATGAT GTAACCTTTC AAGAATTAAG TGGTCGTTCT	2040
GTGGTAAAAA CAGAAGTCGC TTCAGACAGA GAACCATTTG ACGTCTGGCA AGATGCTTGT	2100
GCAAGCTATT CTGAAACAGA ATTAAATATC CAAATTAACG AAGACACGTT TGTTACTTTG	2160

AATCGTCACT TTGTTGTTTCG CATTGACGTG AAAGAAGTTG ACGGTCCTGT CGACAAACAA 2220  
 GTTCGTCGTC GCGATGAATT GATGAACGTG GTCAATACGC TTTCTAATAT GGGCTTATAA 2280  
 AAAATACCAC CAAGCATTTT ACCTTGGTG GTATTTTTTT TAATTAGAAT AAAAGAAACA 2340  
 AGCCAACGGC GCATAGGAAA GAAAGAACTG CGGCACTTAA AAAGATTGTC CGCATGGATG 2400  
 CAGTTTTACT GCTTCTTTTG CCGTGTTTTT TCTGACGTTT TAAACGTTTT TGAAGCTGAC 2460  
 GGGTAATTAC TGAGAAAATC AGCCCGAAAA GGAATAAAC ACTTGTTGCA ATCGATTGGA 2520  
 GAATAAAATG ATTCGGTTCT GCAACGGTGG TATCCCCAGA ATGTTGTTT GATTCTTCCg 2580  
 CAGCTTCCGC TGTTGTTTGT TTTGGCTTAA CTTCTTGGTA AGCCACTTTG GGTGCTTGAA 2640  
 CCGTACTATT TAACAATGTC AAGGGCGTGT CGACTTGAC TGGCCCTGCT ACTAATTTAA 2700  
 AAGTAGGCTG ACTGCCtTTG GGAACGACAC CGTATAAATC GTTAGCCAAA GTAACTTTTT 2760  
 TGTGATCAAT GGTGTGTTT CTTTGTCTA AAAGTTTTTT GCGTTCGTAT GTCGTATAGG 2820  
 CATAGTCCAG CAAAGCATTC CCAGCCACAT GCCGTTTCGT TTCACCATCT TGGTTTTGCC 2880  
 AATTACCAAC GCCTAAAATG ACTTCAATTA AACGGAAATC TCCTTGTTTG ACAGTCGCAA 2940  
 TATAGTTAAA GCCGCCTGTT GGAAGTAGAG CTGTTTTTAA ACCATCCACG CCTTGGTAGC 3000  
 CATACTTAGC TCCTGGTAAA GAATAATTGT ACGCTTCAA CGTTTCCTCG TAAGGCGTGC 3060  
 CGACCATTGT TTTCACCACT GGTATTATGG TGATTTCTAA AATTTCTGGA TAATTTTTCA 3120  
 GAAAATGATA GGTTAGTGAA GCTAAATCTC TAGCTGTAGA GACGTTATCG CCATTTGGGT 3180  
 CAATTCCTTG CATTTGATAC AGGCCGTTAA AGGCACTTGC TTGCGCTCCA CTGCAGTTGT 3240  
 AATAGGTAGT ATTTGTCATC CCTAGTTCAG CCGCTTTTTG ATTCATTAAA TGAACAAAGT 3300  
 CAGTAGGCTG GTTCCCTGAA ATTAAGTTTG CCAACATGAG AGTCGCCACA TTAGAAGAGG 3360  
 GGACAGCAAT CATTTTAAAC AGTTCACGGA CTGGATAAGC AACACCTAAC GTAATTTTGT 3420  
 TATTACTAAT GGCATATATT TTAGAAATAT CGACATCTTT TTGCGTAGCA GTCACAGTCG 3480  
 TATCCATTGT AAATTTTCCT TGCTCCATTG CTTCAAAGGC CAAGTACATG GTCATCACTT 3540  
 TGGCAATACT GGCAGGATTC CACGCTAAAT CTGGTTGCTG CGACCAGAGA ATTTCTCCAG 3600  
 AATCAGCATC AACAACGATG GCTGCTTTGG GAATGCCATC CGCTTTCAA GTCGTCCCCA 3660  
 TTTTTTTAGC AATCGCGGTA ATATCTTCTT CTGCGTGAGT CACAGTAGGT TGGGCAAAAA 3720  
 AGGCCCCACa ACTGACAACT AAGCCTAGCA GAAACAAT 3758

## (2) INFORMATION FOR SEQ ID NO: 44:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1359 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

TCCTCTAATC CTGATGCATT aGCTGGCGTG GATGCACAAC GTGTAGCAGC TTTCCAAGCG

60

GCTAATGGGA AAGCATTGGT CAATCTACGA AAAGCAACGC AAGCCAACAA AGTAAGTTGG 120  
 ACTGTTGTAG CTGCTGCTAG TGAAGGCTGG GCAGCTAAAG TCTTTCCAGA ATTAGCAACA 180  
 AGTGAAGAAC AAGTTGATGC ACTTTGGAAT GAAATTTTCA AAACAACACG AATTTATGAA 240  
 GAAAACCCAG TCATTGCTTG GGATATCCAT GATAAAAAAT TACAAGAAAA AGCGGCTGAA 300  
 TTAAACGAAC AACAATTTAC TGCTTTACAC TACACAGCTC CTGGGACAGA TTTAACGATT 360  
 GGTTCGCCA AAAACCATT ATGGGAAGGC GCTGGTAGTT ATAACGCTCG TGGGGAAGAA 420  
 TTCATGGCTA ATATGCCAAC AGAAGAAGTT TTCACCGCCC CAGATAGTCG CCGCGTTGAT 480  
 GGTATGTTT CTAGTACAAA ACCACTAAGT TATGCTGGCA CAATCATTTT GGAATGAAA 540  
 TTTACATTTA AAGATGGAAG AGTTGTTGAC TTCTCAGCTG AACAAAGCGA AGAAGCATTG 600  
 AAAAATCTCT TAGCTATTGA TGAAGGTGCC AAACACTTAG GAGAAGTTGC TTTAGTACCA 660  
 GACCCTTCAC CAATTTTACA ATCTGGTTTG ATTTTATA ATACATTGTT TGATGAAAA 720  
 GCCTCTAATC ATTTAGCCTT TGGTTCGGCC TATGCCTTCA ACTTACAAGG TGGCACGGAG 780  
 ATGAGCGAAG AAGAATTAGC TGAAGCTGGT TTAAATCGTA GCCAAACACA TGTGGACTTT 840  
 ATGGTAGGAT CTGACAAAAT GAATATTGAC GGCATTAAGG AAGACGGGAC GATTGTTCCA 900  
 GTTTTGTAGAA ATGGCGATTG GGCATAGTCC TACCAAATGT TGAAACCAAA GTTTTGTAGCT 960  
 CCAAGAAATA AGAAGAACT GCTTTTATTC AGTAGTTCT TTTAAATTT AGAAAGTGAG 1020  
 GCCTAAGTTG CAGCGACTTA GGCCTCACTT TCTTATGCTA AAAAATCAAT TGCATTCGAA 1080  
 GGACCTTTAA CGTTATAATC AAATTGAAAA CGGATTCTTT TGCACGATAT AAGTTTACTT 1140  
 TTAAATAAAA GAATTTTTTT CAAAAAGTA GAATTGGAGA ATTAACATGG GATTATTTGA 1200  
 TGGATTATTA GGAAATGCAA CACAAATAA TAATGAAACA GCTGAaAAAG AaTTACGGGA 1260  
 TGTTTTAATC CCTAACGrAA AAGTAGATAT GGCCTTCACT TTAGTAAGAG ATTTAATTGT 1320  
 CTTTACAGAT AAACGCTTAA tTTTrGtCGA TAACCAGGA 1359

## (2) INFORMATION FOR SEQ ID NO: 45:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15614 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

ATTAAAnATAN ACCAGGTTTT AGATAGTATT TATTCAAGTT TACCTCnAAG GTATTAATCT 60  
 GAAACCCkGT TATtTtAtTa GAGAGTTCyT TTTCaAATAT ATGTAAATrG cTAACaACrG 120  
 GGTTCGTGTTT GTTGTGGAC tATTTTTTTT GATCcCCAmC CGTTCTTGAA AAAAAAcAAA 180  
 AAAATAAACA GTGyTCGCTT GATTAAATCGA ACGTTTGTTC GTATAATGAT TTTGCaAGTG 240  
 AGGTGATGTT TGTGAAATTT GATTATGAAA AAGAACCTAG TCGCGATATT CTTTGTATTG 300

ATTGCAAAAG TTTTACGCT AGTGTAGAAT GCGCAGAACG TGGGTTAGAT CCTTTGAAAA	360
CGATGTTGGT AGTGATGAGT AATGCCGAAA ATGCGGGGGG ATTAGTTTTC TCAGCATCGC	420
CAATGGCAAA AAAAGTTTTA GGTGTTTCAA ATGTTACGCG TAAAAATCAA GTGCCCCGCTC	480
ATCCTGATTT ATATGTGGTT CCGCCACGAA TGAATTTGTA CATGAAAAAA AATCAAGAAA	540
TCAATAATTT ATATAAAAAG TTTGTCGCAG ATGAAGATCA CTCTGTGTTT AGTGTAGATG	600
AATCATTTCT TGATGTAAGT GCTTCTTTAA GTTATTTTCA TTGTGAAACT GCCTATAAAA	660
TGGCTCGGAT TATCCAAAGA GTCGTTTATA ATCATGTGGG TATTTATGTG ACCGTTGGGA	720
TAGGGGACAA TCCGTTACTG GCTAAATTGG CGTTGGACAA TGGCGCCAAA CATTGCCAG	780
ATTTTATTGC AGAATGGCGG TATGATCGTG TACCAGATAC TGTCTGGCAG TTGCCTTCTT	840
TGACAGATTT TTGTGGGATT GGTCGCAGAA TGGCTAAACG ACTGAATCGT TTAGGAATTG	900
ATTCTGTATA TGAGTTAGCG CAACGAaTCC GCATCTTTTA CAAGAAACAT TTGGCATGAT	960
GGGGCTTCAA TTGTATGCTC ATTCATGGGG GATTGATCGT ACCTTTTTAG GGAAAAAGC	1020
ACCACATAAA GCTGAAAAAT CATTGGGAA TAGTCAGATT TTACCGAGAG ATTATGCGCG	1080
AAGAGATCAG ATTGAAGTAG TCTTAAAGGA ACTGACAGAA CAAGTGGCTG CTCGTTTAAAG	1140
AAAAGCGCAT TGTCAAACGG AATGTATCAC AGTCTACGTT GGTATTCCA AAGGTCAAAT	1200
TGACCGTGAA GGACGAACTG GTTGGCGAAA ACAACAAACG ATTCCTGCAA CAAACAATAC	1260
AAAAGTTTTC ATCACCTATG TATTAGCCCT TTTTAGAGAA CATTATCTAG CGGGAACCGA	1320
TGTCCGGCAA TTAGGGCTTT CATATGGGAA ATTAGTTGG AATGAATCTT TACAATTAGA	1380
TTTATTTTCG GAACCAGAAG AGCAAATTAG TGAAATGGAG TTAAACTATC TGATTGATAA	1440
AATCCGACAA AAATTTGGCT TTCAAGCTTT AATTCACACC TCTTCTTGT TGGAAGGCGC	1500
GACGGCAATT CATCGTTCAG GATTAGtCGG TGGACATGCT GGAGGAAATG TCGGTTTAGG	1560
AGGATAACAT CATGAAACGA ACAAAAAAAG AmTTTACACC TTATAATGAG TACCATGaTC	1620
GGCCCTTTgA ACTGAAATGG GCGACCGCTT TTGCATTGGG GGAGTTACAA GAAGGAATCG	1680
AAGCTTCTGA CGAATATTCT AAGAGAGCGT TTGAACGCTT ACCGCAACAG ACAACAGAAG	1740
AAATTGAAGA ATACTTAGAG CAGTCTTTAA AAAATAACCA AGTTTtagAa ATCCAATTGA	1800
ATACATTGGA TGATTTTGGC CGTGTTCAAC CGCCAATTGT TGGTGTTTTT GAAGGGATGA	1860
GTTCAATGGA GTATGTAAA GTTGGACATA CACATATTTT ATTTGAAGAG ATTCGGCATA	1920
TTAAAAGTCA TGAGTTTACA AAGTGGTCGA GCCTTGATTA AAAAATATT GTTATACGTG	1980
CGTTTTGTGA AAACATATGA GATAATGTTT ATAGTAATA GAAATAAAAA TTATGGAGGT	2040
GGCTAGGATG GGACATCATT TAAAAGGACA CCATGGCAAA AAAACAAACG CAAATGGCGG	2100
CTTTTTAACA AAGTTATTGG CAGGTGGCGC CTTAGTGGGC GTAGGAAAAA AGTTATACGA	2160
CAATCGTGAA AAAATTAAAG AGCTGTTAAG TGACGACAAA GAAAAAGGCA ACGACACTGA	2220
AAACAAATAA AGGACTGGAT TGACGGAATG GTAAAAGATG CAGCGTATAT TATTGAAGAG	2280

ATTACCACGA TTTTAGAAGG AACCTATTTT GTGATAACCA AGAAAGCGAA TAAGCGCCGC	2340
ATGTGTTTAC ATGTTTCAGA TACCGCGAAA GAATTTTAC GTTGTCGACT GGAGTATGAA	2400
aCATTCTTAG AGCAACAGTT TGATTATGCC aCAATGTTTT TAGAAACATG TACGGAAGAT	2460
GTTTCTGTTG AAATTCAAAC GGGACTACTT GAAGAAACAA ATTATCAAGC ATTAACCGTG	2520
GAAAAATTAC AACAGATAGC CACTAAAATG TAACACAGCT GTTGTGTTAG TTTTGATTCT	2580
TAATTTGGTT AAGGAGGCGG ATAAAATGTC AGTATATCAA CATATTTTGG TTGCTCTTGA	2640
TGGCTCAGAT CAATCTGAAA AAGCTTTTCA TGAAGCTGTC AGGATTGCCA AAGAGGAACA	2700
AGCCACTCTT TATTTAGCTA CAATTATTAA TGATGCTGAA TTTACGACGA GTCCATTTTC	2760
TTTTGAAGAA TTGTATGATT TAGAAAAACA CAAGTCAGAA GAGATGCTCA CTGGAAAGGC	2820
AAAACAAGCG AGTGAAATTG GTGTAAAAAC AGTCAAAAAA ATTGTTGAGC TTGGGAGCCC	2880
AAACGTTAT CTTGCCAACA CGATTTCTGA AAATTATGCC ATAGATTTAA TTGTGCTAGG	2940
TGCCACGGGA AGAGGTGCTA TTCAACGAAC CTTGATTGGT TCTACTACAG ATTATGTGGT	3000
TAACCATGCT TTGTGCAATG TCTTAGTCGT TCGCTAATTG TGAGTAAATA ATCAAGGAAA	3060
CAATGTTTTA TCGGGACGTT TTTTAAATT AAACGAGAGG TGAGTTTATG AAAACAAAAA	3120
TCGGAAAAAC AGTTATCTTG TCAGCATTTT TATTCACAAG TTCCTTTTA CTGAGTGGTT	3180
GTACCTCGGC TGGCGAAGAG ATGGAAAAA CAATTGATCG ACAGAAAGAA AAAGTCGATA	3240
AAACGGTCGA TAAGCAGAAA CATAAAAATG AAAATTCCAT GGAAAGTTAC GACGAAAAAG	3300
TTGACCGTTC TTTAGATAGT CAAGAAGACA AAATCGATAC TACTGAGTAA CGGGAAAGCT	3360
TTCAACTATC GGCTAGTAAA TAAATAATAT ACAAAGGGA GCATCCACTG CAATTATTTT	3420
TCACAGCGGA TCCTCTCTTT TGTGGTTAAA AAGAAAAAGT TAACTGCTCC CCAACAGTTA	3480
ACTTTTAATC ATGTTGCACA TAGTTAAAC ACTCTATATA TCATTTATAT CATTTGAAAC	3540
GGGTAGTGTA AAAAGAAATC TCTTGAAAAG AGTCTCCATT TATAGGGAAG ATTTCCATA	3600
AAAAAGAGAA ACGTATAATC ACAGAGAGTT AGAGAAAGCA AAGGGAAATA AGTATGTAA	3660
AATAAGAAGC AAGAAAGAAT ACAAGTATGC AGAAAATAAC AACTAATTAG GTGAAAACAA	3720
TGGAAATACA TTTTGAAAAA GTAACAAGCG ATAATCGTAA AGCAGTGGAA AACCTACAAG	3780
TTTTCGCTGA ACAACAAGCC TTCATTGAAT CAATGGCAGA AAATCTAAAA GAATCAGACC	3840
AATTTCCAGA ATGGGAATCG GCAGGTATTT ATGACGGCAA TCAATTAATC GGCTATGCCA	3900
TGTATGGACG TTGGCAAGAT GGGCGCGTTT GGCTAGACCG CTTTTTAATT GATCAACGGT	3960
TCCAAGGACA AGGTTATGGG AAAGCAGCCT GTCGACTTTT AATGTTAAAG CTAATAGAAA	4020
AATACCAAAC AAACAAGCTA TACTTGAGTG TTTATGATAC GAACAGCTCa GCAATTCGGT	4080
TGTATCAGCA ACTCGGTTTT GTATTTAATG GTGAATTAGA TACGAATGGT GAGCGGGTAA	4140
TGGAATGGAC ACATCAGAAT AAGTAGATAG AACAGGTGAG ACAAAGCCA TAGCGCATTT	4200
TGTTTCATCT TTTTCTTTTA ATTGGCTTCA TTTGTTTACT TTTAGTAAGT AAAATGGTAA	4260

ATTTAGTTCA ATCTGTTTGG AAAAAAGAGT TACTGGAGGG TAAGGATGAA AATAGAATTT	4320
TTCCATGATG TTATTTGTAG TTTTGTGTTT CCAATGTCTG ACCGCATGCA TGAGATTCAA	4380
CAAGAATTTTCTGAAATTGA ATTGATTCAT CGGTCGTTTG CGTTAGGATG GTCAGCAAAC	4440
GATTTTGAAA CAATGTTTGG CTCACGAGCA GCAGTAAAAG AAGAAGTTTT GACGCATTGG	4500
GTGCATGCCA ATCAAAACGA TTCAAAGCAT CGTTTTAATA TTGAAGGGAT GAGGAAGCAA	4560
CGATTTGATT TTCCaACATC AAAGAATGTA TTATTAGCAG CAAAAGCAGC TGGATACATT	4620
GGCAACCAAG ACACCTATTG GTTACTGTTT GATAAGTTAC AAGAAGGTTT ATTTGTGAGA	4680
AGTTTAAATA TTGAAGAGCC AGAAGTAATT GAGAAGCTAG TAAAAGAGAC AACGATTGAT	4740
TTTGCTTTAT GGAAAGAAGC TGTTGCTTCC GAGGCTGTTT GGACGGCGGT TCAAGAAGAT	4800
TTTGCGCTAG CTTCTGCTTA TGGTTTACAA GGAGTTCCCG CACTAATCAT TAATCAAAAA	4860
TATCTAATTA ATGGTGCAGT ACCGAAACAG CAAATCAGCC AAACCTATTCA AAAAATTCTC	4920
GCAGAAGAAA AACACAGCA ACCGTTAGTT TCATTAAGCTC CTTCTTCTGA AGAGTCAGCT	4980
TCGTGTGAGT TTAAGGAAGG CACATGGACG TGTCATTAG CAGCTTAAAA TAACGTTATT	5040
CTTTGACAAA ACGAAAGACT TCTGTAATAG TTAATGCGGG GGTGATAATA TGTTTAACGC	5100
ATATAGAGAA TACTGGAATA ATATTACAAC AATGAATGCG TCACAACaCG AGCACAATAT	5160
TGGTGGCCAC AATTAATTAA TTATTTGGTG TTAGGGATTT ATTCAGCAAT TACGGGCGTT	5220
TACAGATATA TTGAAGTAAC ACCCAATGAT GGAACAATCA TTAAAGAATG GAATACTGTG	5280
ACCATTATAT TTTTCTTGCT GACTGCTTTA ATTTGGTTGG CAACTTTTAC TGTTGAGCA	5340
AGAAGATTGC ATGATCGCGA CCACAGTAAC TGGTGGATTT TATTCTATTT ACTTCCTTTT	5400
ATCGGTACTT TAGTGATCTT CATTACATTA ATTTTGCCAA GTAAAAGTCA CACAAGATGG	5460
CCAGTGAATC AATCAGAAAT TTAATTTGCT GGTTCGTTT AGCTGATTAT CTTAAGAGTG	5520
AAAATAGGAC ATAAGTCATT TCGACTTATG TCCTATTTTC TTTTTATTT CAGAAAAATA	5580
AAAATATCTT CTGTATGATG GCACCTCACA TTACGCGGTT CTTGATAAAG GGGCTATCTT	5640
TTTGGTAGAT AGGGAAAACC ATTAAGACGT TAAAAGTTTA ATAAATCGAG CAATTCTCTT	5700
TTGTTAGAAA ATAAGATAGC ATCATTAATT AAAATACTTG GAACACTGGT TTCAATAATT	5760
GGTTGATTGG TAATAATCAA GTCAAATTCT TCGAGTCGAT GGTACCTAA TTGTTCTTCC	5820
GTAATTGCTA ACGTATTTGA ACGAATCCGA TGCAAAGCTA AAAtATCGTG GaCATAGTTT	5880
CCAATATATT GACTATGATT CAGGCCCATG TCGCTAATAA TCAAATCTT TTTTGAGTCG	5940
AAACTCTGGT TAATTTCTGG AAAAGTATTA AGCAACCAAA AAATATAAGA AGATGAATAG	6000
TTTAATAGAT TTGTTTCTAA AAGAGCGGAA AAGAAATGCA ATTCTGCTTT TAATCGTGCA	6060
TATAAGTGAA GCTGAGTAAA ATGAATTTGT CGAGTGAAAA AGGTGGCCCT TTGTGTCAAT	6120
TCGGAGGTGT CAAAAGGCAC TAATTGAGAA AACAAATAAA GACAGGTAA CAAGTGAACA	6180
GCGAAGTCGA CCACTTCTTC TGTTAATTGT TCAATCTGTG AAAAAGCATT AAATTTTAAA	6240

TTTTCACGTA TTCTTTGACT GATGACTTCT TGGTCTTCGG AGTAGTATTG CTGAAAAGCG	6300
CCTCTTTGAA AATATCGGAG AACTTCCCGA TAGGTTTCAC GGCTGAGTCC AGTGAAATGA	6360
CGACTGAGTA AAAGATAATC TTCCATTTTG ATTGGACTCT TTTTATCTGT TTTTTTCTGG	6420
AAACGTTGTT GAAATCCTTG ATGCTCTCTT AATAAATCAA CAATAAAATA CATAATAAAG	6480
AAATCGTGTT CATAACTGTT ATCCATaCTA ATAATATTTA ATTGGCATTG ATTACGACGC	6540
ATTAATTGTT TGATTTCCTG CTTATCTAAT TGATAAGGCA ATTCATAATG ACCATAATAC	6600
ATTAAAAAGA AAAAGGTTGT AAAAATACGT AAAGGCAATT CTTCTGTTTG GCTAGTTAAT	6660
TGATAGCCTT GATCAACGAT AACGCGAAGA GAGTAAGTTT CCAAGACCTG GTTACACTGC	6720
TTCAATTTTC GAGCAAAAGA AGAGCGACTA ATATTGATGG TATCGCAATA ATATTGATG	6780
GTTTTGCCAG GATTTTGGAT AACTCATGG AGCAATTGGG TGGTCGTTGA CGTTTTTGCC	6840
ATATCCGTCA GCACGTTTAA ATATAAGGAC ACATTAAAGT TTTCTAAACG ATAGCCTAAT	6900
TTATGCGATG TTTTAATGGT TAACTTTTCG CCCCACGCT GTTTAATCTC CTCAATATAA	6960
CGAATAAAAG TAGACTCAGA ACTACCAATA TATTTTGCAA GTTCTTTTTT AGTCATCCAT	7020
TTATCAGAAA TGGATAAGAC ACGAATGATT TCTTCCATTT TCCGATAATT TTTTCTAAA	7080
AACCTAATCA TATTGTTTCC TCCCAAGTCC TTTACTTTAA TGAAACAATC ATCGAAACAA	7140
TAACGCAATT TCAGAAATGA AACCGCTTGG TATGAAAGCT GGTGTGATTT AATGAGTTCC	7200
TTCTCTGTAT AATAATTATC AGATTAGAAT AATGAAATTT TACGTGCTAG GCAATTCTaA	7260
TTTAACTGkT AAwTcmTaAA AGgAAAGGAG AGAACAGACA TTAAATTACT AAAGAmATTT	7320
GGGGTTTTCG GCGTAGTCTT TTTACTTGTC TCATCCTATT TCATACCGCT CATTGGCTAy	7380
GCAGAAACAG CCAAAGAAGT TGAAATTACA TCCGCTCAAA TGATAACAGA TGAGAATGAT	7440
AAAACGAATA TTAATATCGA GTTAAATCTT CTCAACCAAA CAGAGCAGCC ATTACAACGA	7500
GAAATTCAAT TGAAAAATGC ACAGTTCATG GATACTGCTG TAATTGAAAA AGACGGATAT	7560
TCTTACCAAG TGACTAATGG TACGCTTTAT CTGACTTTGG ACGCACAAGT AAAAAAGCCG	7620
GTACAGCTTT CGTTAGCTGT TGAGCAAAGT TCGCTTCAAA CAGCTCAGCC ACCTAAGTTA	7680
TTGTATGAAA ACAACGAATa TGATGTTTCA GTTACTTCTG AAAAAATAAC AGTAGAGGAT	7740
TCTGCTAAAG AATCAACTGA ACCAGAAAAA ATAACTGTAC CAGAAAATAC GAAAGAAACT	7800
AACAAAAATG ATTTCGGCTCC AGAAAAACA GAACAGCCGA CCGCAACAGA AGAGGTAACC	7860
AATCCATTTG CAGAAGCAAG AATGGCGCCA gCTACTTTGA GAGCGAATCT GGCACTGCCT	7920
TTAATGCAC CACAATACAC GACGGATAAT TCTGGGACTT ATCCGACAGC TAATTGGCAG	7980
CCCACAGGCA ATCAAAATGT GTTAAACCAT CAAGGGAATA AAGACGGTAG TGCACAATGG	8040
GACGGCCAAA .CGAGTTGGAA TGGGGACCCT ACTAATCGCA CAAATTCTTA TATTGAGTAT	8100
GGCGGTACAG GAGACCAAGC CGATTATGCC ATCCGAAAAT ATGCTAGAGA AACACAACA	8160
CCAGGGCTTT TTGATGTATA TCTTAATGTG CGTGGGAATG TTCAGAAAGA AATCACGCCA	8220

TTGGATTGG	TCTTAGTCGT	TGACTGGTCC	GGTAGTATGA	ATGAAAACAA	TCGGATTGGT	8280
GAAGTTCAAA	AAGGAGTGAA	CCGTTTTGTT	GATACATTGG	CAGATAGCGG	TATTACCAAT	8340
AACATCAACA	TGGGCTATGT	TGGCTACTCA	AGTGACGGTT	ATAATAACAA	CGCCATTCAA	8400
ATGGGGCCGT	TTGATACAGT	CAAAAATCCA	ATTAAAAATA	TTACGCCAAG	TAGCACTAGA	8460
GGAGGAACTT	TCACTCAAAA	AGCATTAAAG	GATGCTGGTG	ATATGTTAGC	AACGCCAAAT	8520
GGACATAAGA	AAGTCATTGT	ACTTTTAACG	GATGGCGTCC	CAACCTTCTC	TTATAAAGTG	8580
AGTCGAGTTC	AAACAGAGGC	GGATGGTCGC	TTTACGGGA	CACAATTTAC	GAATCGACAA	8640
GATCAACCAG	GTAGCACTTC	TTATATCTCT	GGTAGCTATA	ATGCGCCAGA	TCAAAACAAT	8700
ATCAATAAAC	GGATTAACAG	TACGTTTATC	GCCACGATAG	GTGAGGCAAT	GGTCTTAAAA	8760
CAACGTGGGa	TTGAAATACA	TGGATTGGGC	ATTCAATTGC	AAAGCGATCC	ACGAGCTAAT	8820
TTATCTAAAC	AACAAGTTGA	AGATAAAATG	CGTGAGATGG	TGTCAGCCGA	TGAAAATGGA	8880
GACCTTTATT	ATGAATCCGC	GGATTATGCA	CCAGACATTT	CTGATTATTT	AGCGAAAAAA	8940
GCCGTTCAGA	TTTCAGGAAC	GGTTGTAAAC	GGAAAAgTAG	TTGATCCAAT	TGCTGAACCT	9000
TTTAAATACG	AGCCAAATAC	ATTATCAATG	AAAAGTGTGG	GTCTGTTC	GGTTCAAACA	9060
TTACCAGAAG	TGTCGCTAAC	AGGCGCTACA	ATTAATAGTA	ATGAGATTTA	TTTGGGTAAA	9120
GGGCAAGAAA	TTCAAATTCA	TTATCAAGTA	CGTATTCAAA	CAGAGTCAGA	AAACTTCAAA	9180
CCTGATTTTT	GGTATCAAAT	GAATGGTCGG	ACAACGTTTC	AGCCATTAGC	CACGGCCCCCT	9240
GAAAAAGTTG	ATTTTGGGGT	TCCTTCGGGA	AAAGCACCTG	GCGTGAAGTT	AAACGTGAAA	9300
AAAATCTGGG	AAGAGTATGA	TCAAGACCCG	ACAAGTCGGC	CAGATAATGT	GATTTATGAA	9360
ATTAGTAGAA	AGCAAGTAC	TGACACAGCC	AACTGGCAAA	CTGGGTATAT	TAAATTATCA	9420
AAACCAGAAA	ATGATACCAG	CAATAGTTGG	GAGCGCAAAA	ATGTAACCCA	ACTTTCCAAA	9480
ACCGCGGATG	AAAGCTATCA	AGAAGTTCTT	GGGCTTCCCC	AATACAACAA	TCAAGGACAA	9540
GCTTTCAATT	ATCAAACAAC	CCGTGAATTA	GCAGTTCCTG	GTTACAGTCA	AGAAAAAATC	9600
GACGATACTA	CTTGGA AAAA	CACGAAGCAG	TTCAAGCCAT	TAGATTTAAA	AGTAATCAAA	9660
AATTCTTCCT	CAGGTGAGAA	AAACTTAGTG	GGAGCCGTCT	TTGAATTGAG	TGGTAAAAAT	9720
GTTCAAACAA	CATTAGTGGA	CAATAAAGAT	GGTAGCTATT	CCTTGCCAAA	AGATGTGCGC	9780
CTACAAAAAG	GGGAACGCTA	TACATTAAC	GAAGTAAAAG	CACCTGCAGG	ACATGAGTTA	9840
GGCAAGAAAA	CGACTTGGCA	AATTGAGGTG	AGTGAGCAAG	GCAAAGTAAG	CATCGATGGA	9900
CAAGAAGTGA	CCACCACAAA	TCAAGTTATT	CCATTGGAAA	TTGAAAATAA	ATTTTCTTCT	9960
TTGCCAATCA	GAATTAGAAA	ATACACCATG	CAAAATGGCA	AACAAGTGAA	CTTAGCAGAG	10020
GCGACTTTTG	CGTTGCAAAG	AAAAATGCT	gCAAGGAAGT	TACCAAAC	TG	10080
AAAAACAGAT	ACTACAGGAT	TGAGCTATTT	TAAAATTAGT	GAACCTGGTG	AGTATCGAAT	10140
GGTGGAACAA	TCAGGACCAT	TAGGCTACGA	CACTCTTGCT	GGAAATTATG	AATTTACTGT	10200



TGATAAATAT GGGAAAATTC ACTATGCAGG CAAAAATATT GAAGAAAATG CGCCAGAATG	10260
GACACTGACA CATCAAATA ATTTGAAACC TTTTGACTTA ACAGTTAATA AAAAAGCCGA	10320
TAATCAGACG CCACTTAAAG GAGCGAAATT CCGTTTAAACA GGACCAGATA CGGATATTGA	10380
ATTACCAAAA GATGGCAAAG AAACGGATAC TTTTGTTTTT GAAAACTTAA AACCAGGGAA	10440
ATATGTTCTA ACAGAAACCT TTACGCCAGA AGGATATCAG GGGTTAAAAG AACCAATCGA	10500
ATTAATAATT CGTGAAGATG GTTCAGTCAC GATAGATGGG GAAAAAGTAG CAGATGTTTT	10560
AATTTCTGGA GAGAAGAATA ATCAAATTAC TTTAGACGTT ACGAACCAAG CAAAGGTTCC	10620
TTTACCTGAA ACTGGTGGCA TAGGACGCTT GTGGTTTTAC TTGATAGCGA TTAGTACATT	10680
CGTGATAGCG GGTGTTTATC TCTTTATTAG ACGACCAGAA GGGAGTGTGT AATCAATGAA	10740
AAACGCACGT TGGTTAAGTA TTTGCGTCAT GCTACTCGCT CTTTTCGGGT TTTCACAGCA	10800
AGCATTAGCA GAGGCATCGC AAGCAAGCGT TCAAGTTACG TTGCACAAAT TATTGTTCCC	10860
TGATGGTCAA TTACCAGAAC AGCAGCAAAA CACAGGGGAA GAGGGAACGC TGCTTCAAAA	10920
TTATCGGGGC TTAAATGACG TCACTTATCA AGTCTATGAT GTGACGGATC CGTTTTATCA	10980
GCTTCGTCTT GAAGGAAAAA CGGTCCAAGA GGCACAGCGT CAATTAGCAG AAACCGGTGC	11040
AACAAATAGA AAACCGATCG CAGAAGATAA AACACAGACA ATAAATGGAG AAGATGGAGT	11100
GGTTTCTTTT TCATTAGCTA GCAAAGATTC GCAGCAACGA GATAAAGCCT ATTTATTTGT	11160
TGAAGCGGAA GCACCAGAAG TGGTAAAGGA AAAAGCTAGC AACCTAGTAG TGATTTTGCC	11220
TGTTCAAGAT CCACAAGGGC AATCGTTAAC GCATATTCAT TTATATCCAA AAAATGAAGA	11280
AAATGCCTAT GACTTACCAC CACTTGAAAA AACGGTACTC GATAAGCAAC AAGGCTTTAA	11340
TCAAGGAGAG CACATTAACT ATCAGTTAAC GACTCAGATT CCAGCGAATA TTTTAGGATA	11400
TCAGGAATTC CGTTTGTGAG ATAAGGCGGA TACAACGTTG AACTTTTAC CAGAATCAAT	11460
TGAGGTAAAA GTGGCTGGAA AAACAGTTAC TACAGGTTAC AACTGACGA CGCAAAAGCA	11520
TGGATTTACG CTTGATTTTT CAATTAAAGA CTTACAAAC TTTGCAAATC AAACAATGAC	11580
TGTGTCGTAT CAAATGCGTT TAGAAAAGAC CGCTGAACCT GACACTGCGA TTAACAACGA	11640
AGGACAATTA GTCACGGACA AACATACCTT GACTAAAAGA GCCACAGTTC GTACAGGCGG	11700
CAAGTCTTTT GTCAAAGTTG ATAGTGAAAA TGCGAAAATC ACCTTGCCAG AGGCTGTTTT	11760
TATCGTCAAA AATCAAGCGG GGAATACCT CAATGAAACA GCAAACGGGT ATCGTTGGCA	11820
AAAAGAAAAA GCATTAGCTA AAAAATTCAC GTCTAATCAA GCCGGTGAAT TTTCAGTTAA	11880
AGGnnTAAA AGATGGCCAG TACTTCTTGG AAGAAATCTC TGCACCAAAA GGTATCTTC	11940
TGAATCAAAC AGAAATTCCT TTTACGGTGG GAAAAATTC TTATGCAACG AACGGACAAC	12000
GAACAGCACC GTTACATGTA ATCAATAAAA AAGTAAAGA GTCAGGCTTC TTACCAAAAA	12060
CAAATGAAGA ACGTTCTATT TGGTTGACGA TTGCAGGCCT GCTAATCATT GGGATGGTAG	12120
TCATTTGGCT ATTTTATCAA AAACAAAAAA GAGGAGAGAG AAAATGAAGC AATTAAAAAA	12180

AGTTTGGTAC	ACCGTTAGTA	CCTTGTTACT	AATTTTGCCA	CTTTTCACAA	GTGTATTAGG	12240
GACAACAAC	GCATTTGCAG	AAGAAAATGG	GGAGAGCGCA	CAGCTCGTGA	TTCACAAAAA	12300
GAAAATGACG	GATTTACCAG	ATCCGCTTAT	TCAAATAGC	GGGAAAGAAA	TGAGCGAGTT	12360
TGATAAATAT	CAAGGACTGG	CAGATGTGAC	GTTTAGTATT	TATAACGTGA	CGAACGAATT	12420
TTACGAGCAA	CGAGCGGCAG	GCGCAAGCGT	TGATGCAGCT	AAACAAGCTG	TCCAAAGTTT	12480
AACTCCTGGG	AAACCTGTTG	CTCAAGGAAC	CACCGATGCA	AATGGGAATG	TCACTGTTCA	12540
GTTACCTAAA	AAACAAAATG	GTAAAGATGC	AGTGTATACC	ATTAAAGAAG	AACCAAAAGA	12600
GGGTGTAGTT	GCTGCTACGA	ATATGGTGGT	GGCGTTCCCA	GTTTACGAAA	TGATCAAGCA	12660
AACAGATGGT	TCCTATAAAT	ATGGAACAGA	AGAATTAGCG	GTTGTTTCATA	TTTATCCTAA	12720
AAATGTGGTA	GCCAATGATG	GTAGTTTACA	TGTGAAAAAA	GTAGGAACTG	CTGAAAATGA	12780
AGGATTAAAT	GGCGCAGAAT	TTGTTATTTT	TAAAAGCGAA	GGCTCACCAG	GCACAGTAAA	12840
ATATATCCAA	GGAGTCAAAG	ATGGATTATA	TACATGGACA	ACGGATAAAG	AACAAGCAAA	12900
ACGCTTTATT	ACTGGGAAAA	GTTATGAAAT	TGGCGAAAAT	GATTTCACAG	AAGCAGAGAA	12960
TGGAACGGGA	GAATTAACAG	TTAAAAATCT	TGAGGTTGGT	TCGTATATTT	TAGAAGAAGT	13020
AAAAGCTCCA	AATAATGCAG	AATTAATTGA	AAATCAAACA	AAAACACCAT	TTACAATTGA	13080
AGCAAACAAT	CAACACCTG	TTGAAAAAAC	AGTCAAAAAT	GATACCTCTA	AAGTTGATAA	13140
AACAACACCA	AGCTTAGATG	GTAAAGATGT	GGCAATTGGC	GAAAAAATTA	AATATCAAAT	13200
TTCTGTAAAT	ATTCCATTGG	GGATTGCAGA	CAAAGAAGGC	GACGCTAATA	AATACGTCAA	13260
ATTCAATTTA	GTTGATAAAC	ATGATGCAGC	CTTAACCTTT	GATAACGTGA	CTTCTGGAGA	13320
GTATGCTTAT	GCGTTATATG	ATGGGGATAC	AGTGATTGCT	CCTGAAAATT	ATCAAGTGAC	13380
TGAACAAGCA	AATGGCTTCA	CTGTCCGCGT	TAATCCAGCG	TATATTCCTA	CGCTAACACC	13440
AGGCGGCACA	CTAAAATTCG	TTTACTTTAT	GCATTTAAAT	GAAAAAGCAG	ATCCTACGAA	13500
AGGCTTTAAA	AATGAGGCGA	ATGTTGATAA	CGGTCATACC	GACGACCAA	CACCACCAAC	13560
TGTTGAAGTT	GTGACAGGTG	GGAAACGTTT	CATTAAAGTC	GATGGCGATG	TGACAGCGAC	13620
ACAAGCCTTG	GCGGGAGCTT	CCTTTGTCGT	CCGTGATCAA	AACAGCGACA	CAGCAAATTA	13680
TTTGAAAATC	GATGAAACAA	CGAAAGCAGC	AACTTGGGTG	AAAACAAAAG	CTGAAGCAAC	13740
TACTTTTACA	ACAACGGCTG	ATGGATTAGT	TGATATCACA	GGGCTTAAAT	ACGGTACCTA	13800
TTATTTAGAA	GAACTGTAG	CTCCTGATGA	TTATGTCTTG	TTAACAAATC	GGATTGAATT	13860
TGTGGTCAAT	GAACAATCAT	ATGGCACAAC	AGAAAACCTA	GTTtCACCAG	AAAAAGTACC	13920
AAACAAACAC	AAAGGTACCT	TACCTTCAAC	AGGTGGCAAA	GGAATCTACG	TTTACTTAGG	13980
AAGTGGCGCA	GTCTTGCTAC	TTATTGCAGG	AGTCTACTTT	GCTAGACGTA	GAAAAGAAAA	14040
TGCTTAATTT	CTAGCATCAC	CGAAGAAATT	TTTAGAAAAA	CAAAGAGCCT	GGGCCAATCA	14100
CTGTCCCAGG	CTCTCATGCT	TTATTTTTAA	GGAGGAAGCA	ATGAAGTCAA	AAAAGAAACG	14160

TCGTATCATT GATGGTTTTA TGATTCTTTT ACTGATTATT GGAATAGGTG CATTGCGTA	14220
TCCTTTTGTT AGCGATGCAT TAAATAACTA TCTGGATCAA CAAATTATCG CTCaTTATCA	14280
AGCAAAAGCA AGCCAAGAAA ACACCAAAGA AATGGCTGAA CTTCAAGAAA AAATGGAAAA	14340
GAAAAACCAA GAATTAGCGA AAAAAGGCAG CAATCCTGGA TTAGATCCTT TTTCTGAAAC	14400
GCAAAAAACA ACGAAAAAAC CAGACAAATC CTATTTTGAA AGTCATACGA TTGGTGTTTT	14460
AACCATTCCA AAAATAAATG TCCGTTTACC AATTTTGTAT AAAACGAATG CATTGCTATT	14520
GGAAAAAGGA AGCTCCTTGT TAGAAGGAAC CTCCTATCCT ACAGGTGGTA CGAATACACA	14580
TGCGGTCATT TCAGGCCATC GTGGTCTCCC TCAAGCCAAA TTATTACAG ATTTGCCAGA	14640
ATTAAAAAAA GGCGATGAAT TTTATATCGA AGTCAATGGG AAGACGCTTG CTTATCAAGT	14700
AGATCAAATA AAAACCGTTG AACCAACTGA TACAAAAGAT TTACACATTG AGTCTGGCCA	14760
AGATCTCGTC ACTTTATTAA CTTGCACACC GTATATGATA AACAGTCATC GGTTATTAGT	14820
TCGAGGACAT CGTATCCCAT ATCAACCAGA AAAAGCAGCA GCGGGGATGA AAAAAGTGGC	14880
ACAACAACAA AATTACTAT TATGGACATT ACTTTTAATT GCCTGTGCGT TAATTATTAG	14940
CGGCTTCATT ATCTGGTACA AGCGACGGAA AAAGACGACC AGAAAACCAA AGTAGTATGA	15000
CGAAAAGGCT AAACATACTA AAAAAAGAG TAAAAAATA GCTTTTCAAT TTTAATCCT	15060
CCTTATCGTG CATAATTGAA CCAGAGAAAC AGAAGTATTA ACGAAATAAC TAAAAGAGCA	15120
AGCCCTGAAT AAAAAGCGAC AAAGGGCCAA TCAATCGACT GTTTAAATTC CTGCCAAGTT	15180
TGGATTTTTC TGTTTTTTTT CGCGCTATCC TCAAGCGTGA GTAAATAATT CAATAGTAAG	15240
AGGAGTAGCA ACACCGTGAA ATCATTGTG GTAAAAAGCA CATGTAAAAA TAGAATGACA	15300
AAGACAACAC GGGATAACAC TCGATTCCGC AAAATTAAAA ATAACCTAGC ACGCATAATA	15360
AACCACCATT TCTTaTCAGA GATAATGAAT CTGTTTTTGT CTACTCTTTA GTTATATCAT	15420
AAAATTCTTA ATAATGAAAA AATGACTCGA GAAATAAATT GAAAAAGTT TTTTTCTTG	15480
AATCATTATT TTCGTAAATA AAGAATAAAC GTGTTACTCT TGGCTTATCA AATTGGAAG	15540
GAGTGTTAAA AATGAAATAT CTGGATATTA TTGCTTTAAT TTTATTGATT GTCGGAGGT	15600
TAAACTGGTT ATTA	15614

## (2) INFORMATION FOR SEQ ID NO: 46:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3169 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

TTCACAAAAA AATGAAAACA CAAATTTAAC TTGGAGGATG TTGAGTAGTG AGCAACGGGA	60
AAAAGATTTT nATTAGTCAT AGTTCAAAGG ATcAAGAATA TGTGGATGCG TTTATCCAAT	120
TATTGAAAAA GTTTGGGTTT CGAACACAGG ATATTTTTTA TAGTTCCACT ATTGAAACAG	180

GTGTACAGCC AGGAGAATTA ATTTTGTGACA CGATTAAAAG GGAGCTAACC AATCAACCTG	240
TCATGCTGTA TTTTATCT GATCATTATT ATCAGAGTAT TCCATGTTTA AACGAAATGG	300
GTGCTTCTTG GATGCTATCA GATAAACACT ATCCAATTGC GTTAAATAAT TTTTCTATGA	360
AAGATATGAA AGGTGTTATT AGTAGTGAGC GTTTGGCGAT TCGGTTTAAT GATAAGACAA	420
GTAATAATGA AATAAACTGc TTATTAAAAA AGTTATCTCA TGACACAGAT GTACAAGCTG	480
AGCCAGATTT TGAATAAAC GTTGAGAAAA ATATTCAGCC ATTCCAAAAT AAAGTACAC	540
AACTGATTCTG GCAAGCAAGT TATTTAAAAA CTGATGAAAA GGGATATTTT GAAACGrTTC	600
tTAGCACTCA TCGTCCTGTG TATGGCACAG CAAAAGGGGT CTATGACTGT TTAAATTAC	660
CTAGCTTAAT TGAGCCTAAA AGTTTAGGTT TAGACACGCT ATCAGAGGAT GAAAGCCATT	720
GGCTATTCTT TTTTCTTACG TGGGGAACCTT TTCAAGAAGG CGAAAAAGTG CGGTTTAAAt	780
TGAAAAAGGa TAAGGcGTAT AACAAATCGGG AATTCAGTGA TATTGGTAAA TGTAATAATA	840
TTTACGTCTC TTATTTAGAG AAGGTAGAGT AATTGGAGGA GGTAATAAAT GAAATGAAT	900
TGGcGTCAGT TGGTTGGTAC ATATCGTATT TCTGATAATG GTAAACAATA TCAAAATGAC	960
TGGTTAAATA AGCAAGTTGA TTTAGCCACA GAACCTTCGTA AATCTTTTGA TAGTGACTAT	1020
CGTAGAGTAA TTAAAAGTGA TGCATTTaGa CGCTTwnCAa GAKAwAACAC cAAGTTTTyC	1080
CTTTwGrAnC GTAATGaTTT TgGTAAGAAC aCGTTTAACA CATAGCTTAG AAGTGGCAAT	1140
GCATAGCAGG GATATACTTA GATTAGTTAT TTCACAGTTG AACGAACGAA ACATTGAGGT	1200
TATGGAATTA TCAGAAATGCT ATCGGTTGTT AGAAACAGCT GCATTAATCC ATGATATCGG	1260
CAATCCCCCT TTTGGACATT TTGGGGAAGA AGCCATTCTG ATTTGGTTTG AAAAAATGG	1320
TCCTTTGTAT GCTTGTGGT CTGATTTTTC GGAGCAACAA CAAAATGATT TTCTTAACCT	1380
TGAAGGAAAC GCTCAAACCA TCCGACTTTT AACAAACTG CATCATGATA ATGGGACTTC	1440
TAAAGCAGGA ATGAACTAA CAGCCACGAC ATTAGATACA GTTATCAAAT ATACGGCATC	1500
TTCAGATCAA TTAAATAAGC AACAGCTTTT GTCTAAAAA GTTGGCTATT TTTACTCGGA	1560
AAAACAAGTG TTTAACGCCA TTAAATTTGC GACAGGAACA CmAAATAAGC GACACCCATT	1620
AGTCtTTTTG TTAGAAGCTT CAGATGATAT CGCCTATACT TTTTCAGATA TCGAAGATGC	1680
CTATAATTAT GGTTTATACA GCTATCaAGA TTTAAAGCaG TTTGTTGATG CGCAAACAGG	1740
AACAAAAAAA TTTTAAATGT TAGAAAAAGA TAAACAGAA GAAGCTACCT TACAAGAATT	1800
TTTGAGAAAG ACACAAAGAG CTGTTTGTCA AAGCGTAGCG AAAAATTTTG CTAATCATT	1860
TGACTTAATT ATGAGAGGTA TCTATCATAA TGAAATTATC ATTGAAGATT GTGATGAAGT	1920
GAAATGCTTT CAGGCATTAA AAAACTTTTC TAGAGTCTAT GTTTTCAAA CAAAACAAT	1980
TTTGGATCAA GAGGTATTGG GCTTCAATAT CATCAATCGT TTAGTAGATG AATTTGTTCC	2040
AGTCGTACTG AAATATGAAA AAGTTTCCAT GAATAAGTAT GAGGAGCGAA TCTTTAATAA	2100
TATTTCGGAA AGCGCGAAAG CGTTGTATAG AAGAGAAGCA AAGAAGCTA CAGAAGCAGA	2160

AAAAGACTAT TATCGTTTGA AAATGGCGGT GGACTTTGTA TGTAATATGA CGGATGGATA	2220
TGCCAAAAAA GTTTATGATA CATTATTCAC TTAAAATTTG AGAAAAAGCT TCTCAGCGGT	2280
TGCCGCCGCT GAAGAAAAC TGTTCATTGC ACAGGCTGAT CGTTTCAGTA AAAATTTTAA	2340
AGATGCTAGT AATTTCACTG AAAATCGGGT ACACTAAAAT TAGCTAGTGT AACTAGCAAA	2400
AAAGGCAACT GGAATAGATA GGTTGTGCTG TCTATGAAGG GCTTTGTTTG TCAGGTGGTG	2460
CTGACATGCA AAGTGCCTTT TTTCTTTTTC TCAACATTTT AAGGACAAAA AACGATTGCA	2520
ACATGTCACA ATCGCTAACC TCGGAGAAAA AGGAATTTTT CTCCAATGGC CATTTCGATA	2580
ACCTGTTGTG TGTTTCGAAA CATTTCAGAAC GGTGGTGCCT TCTTTCAGTG CCTAATTATC	2640
CAATTCGTTT CTTGTTCCCTC TATTTACAGT ATACTGAATG AAAGCCTATT CTGCAATGAT	2700
AGAATCAGGT TCCTGTGATT AAATAACATA AAAAATGGGC TCCTTATATA AAGAAAAATA	2760
ACAAAAGAAA TTTTGAATAT TGACGTTTTT TTTAGGTACG TTATGATAAG CGAGAGGTGA	2820
TTTTATGTTA AAAACGCAAA TTGTTGATTA TTTACGCAAT CAAACAGCTT TTTTGAACC	2880
ATCTTTAGTC AGTGAAATTT TCACAGCAAG CAATATTGCA ACGACATTTT CTATCAAACG	2940
AAATACAGCT AGTCACTATT TAAATCAGTT AAACGAAGAA GGAATTCTAG TTAAGATCAA	3000
TACACGACCG GTTTACTTTT TTCATAAAgA AGCGTTTCAG CAACAAAAC TCTTTTAAA	3060
AAGGAAGTGT TTATCCAAGT TTCCAAGAAA TGATCGATGA ACAGCCTGTT TTTGGATCGC	3120
CAAAGnGGAn TTTTnTTCAG TCGGGTCATn GGGTATCGCG GGTAGTTTA	3169

## (2) INFORMATION FOR SEQ ID NO: 47:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17087 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

GTCTATACAA GTTCACAAGG AGGCGCCAAA CATGACCATT TTAGAATTAA CATTGCAAC	60
GAACCAACTA ACGACGACTT TTTTAAACCA TAAATTGGAA CCGCTTCACC AAAAAACATT	120
CCCTTTTGAA GCCTTGACTA TCGAAGAAAC GGTTCCTGAA CTGTGCCAAT TGATTTTGGC	180
AGAGGCGGCA CCGATTCTGG GGGGGATTGT GGCTGATTTC CCAGCGGACT TCTTTTCGCC	240
AGAGGAAACG TTGGCGCTGG TTCCGCAAGC GGTGATTCAA GCATTGCGCG AACAGTTGAA	300
CACGCCACTT TTAACGGCAG CCGAACAAGC GCGCGCACCG AATCTGCTAG AAGCATTGTA	360
AGAAAAACGG cAGTATCTGG CTTGGCATT TGGACTATTAT GGCTACGTGC CTGGGAAAAA	420
TGAATATGCG GTGGAGTCCC TGTTAACTGT CGGCAATGGC TTTTtaggCT TGCGCGGCAC	480
GACACCTGAG ATGACGATTT CTGATGATCA TTATCCAGCG ACCTACATTG CGGGGCTTTA	540
TAATACAGCG GCTTCGGAAG TTGCGGGTCA AGTAGTGGA AATGAAGATT TCGTCAATGC	600

TCCTGACAAC CAACACATCG CTTTAAAAAT TGGCGATGCC ACTGACTGGT TAACGATTTC	660
TCCAGACACG CTACAACAAT TACATCGTCA ATTGAACTTG AAAACGGGGT TGTTCGTCGC	720
AGAAATGATT TTGAAAGATG CGGACAATCA GCAAATTAAA TTAACGACGA AAAAAATCGC	780
CAATATGGCA CAGCCGAATG ACTATCATCT GCAATATACG TTTGAGCCGT TAAACTTTTC	840
AGCCCCGATT ACTTTGAAAA CCGTCACGGA TGGCAGTGTC TACAACTATA ACGTGGCGCG	900
GTATCGCAAC TTAACAGCGA AACATTTTCA GGTGACAGCA TTAAGCGCGC AGGAAAACAA	960
AACCGTGATT GAGGTTTGTA CCAATCAATC CAACTTAAGC GTCCGAGAAA CGTCTTTGAT	1020
AACCGGTGAT TTTTTTGAAA AAGAAGCCAT TACGATTCAA GAAGAAGCGG AGAAAATCGC	1080
CCAAGTGGTG ACCGTGATGG CGCACCAAGG AACGCGCTAT ACGCTAGAAA AACAAGTCTT	1140
TGTACAAGCC AGCCACGCCG AACAAAGTTG GCAGGTACCG TTTACGCCGA AGGACTCTTT	1200
TGCGGCCGCC GCTCAGGAAA GCGCCCGCGC TTGGCAAAC CTTTGGCAAC AAGCCAACAT	1260
TACCGTGACA GGCGATTTGA TGTCTCAAAA ACTTTTACGG ATTCACAGTT ATCACTTGCT	1320
GGCTTCGGCT TCGCCTTTTA GTAACCAAGC GCAAGCATTG GACGTCTCGA TTACCGCCCG	1380
AGGCTTGCAT GGTGAAGCGT ATCGTGGACA TATTTTCTGG GATGAAATTT TTATTTTACC	1440
TTTTTATATT CAACATTATC CTGACACAGC GAAACAATTG TTGCTGTACC GCTACCACCG	1500
ATTAGAGAAG GCGAAGGAAA ATGCGGCGGC CAGCCAGTAT CGTGGCGCAA TGTATCCGTG	1560
GCAATCAGGG CGTGATGGGC GTGAAACCAC GCAAAAACCTT CACTTAAATC CGTTGAATGG	1620
ACATTGGGGC GAAGATCATA GTATTTTGCA ACGGCATGTT TCCTTGGCGA TTGCTTACAA	1680
TGCTTGGTTG TATTGGCATA GTACACAGGA TCATGAATTT ATGAAACAAT ATGGCGGCGA	1740
AATGTTGTTA GAAATCGCTC AATTTTGGAA CAGCGCTGCG ACGTTAGATG ACGCAACAGG	1800
ACGTTCTTTT ATTGATAAAG TGATGGGACC TGATGAGTTT CACGAGGGCT ATCCAGACCA	1860
AGCGGAAAGC GGCTTGAAAA ACAACGCCTA TACGAATTTG ATGGTGGTTT GGCTATTCGA	1920
AGAATTAACC AACATTTTGG CACTTTTCTC TGAGGAAGAA CAGGCACAAC TTTTGGCAA	1980
AACGCAGACC ACTTCGGCTG ATTTAGCCCG CATGCAGCAA ATTCAAAATT CACTGGAAAT	2040
TGAAGTTAAT TCTGACGGTA TTATTGCGCA ATATGAAGGG TATTTCGGCT TAAAAGAGAT	2100
TGACTGGACA GCGATGAAAG AGAAATACGG CAATATTTAC CGCATGGATC GAATTTTAAA	2160
AGCCGAAGGA GAATCCCCG ATGATTACAA AGTGGCGAAA CAAGCCGACA CGTTAATGTT	2220
GTTCTACAAC TTAGACAAAA CTCGGGTGGA TCAAATTTTA GAAGATTTAG GGTATCAATT	2280
GCCTGCGGAT TATTTGGAAA AAAATCTTTT GTATTATCTA AAACGGACGT CTCATGGTTC	2340
GACTTTGTCA CGAATTGTTC ATGCCCAATT GGCAGAAATG GCGCAATTCC ATGAGCTATC	2400
TTGGCAACTT TATCAAGAAG CGTTGTATTC CGATTATCGT GACATTCAAG GTGGCACGAC	2460
CGCTGAAGGG ATTCATACAG GGGTCATGGC TGCGACGATC CACGTTACTT TAGCGACGTA	2520
TGCAGGCGTT GACACGCGCC AGAACGAGTT ATCAATTTGT CCAAACCTTAC CTGAACATTG	2580

GCAAGCGCTG	GCCTTCCAGT	TTATTTCATCA	AGGCGTGACG	TATCAATTTT	CATTAACACA	2640
AACCAAGTGC	ACGATTACAG	CAGATAGAGA	CACACAATTA	TTGGTTCAGG	GCGCGTTGAT	2700
TCCGTTAACC	GCCGAACGAC	CAAAAGAAGT	ACATTATCAA	TAAAGGAGCG	AACCACAATG	2760
ACGATTGGAT	TTATTTTTGA	TTTGATGGT	GTCATTACTG	ACACCGCCAA	ATTTTCATTAT	2820
CAAGCTTGGA	AAGCGTTGGC	CGATTGCTA	GGTATCCCCA	TTGACGAAAC	CTTCAACGAA	2880
ACATTAAAAG	GCATTAGTCG	GATGGATTCC	TTGGACCGTA	TTTTAGTCCA	TGGCCACCGC	2940
GAAAAATGCGT	TTACTCCAGC	AGAAAAAGAA	GCCTTAGCTC	AGCAAAAAAA	TGAnCaCTAC	3000
GTTCAATTAT	TAGAACAATT	AACAACAGAA	GATGTTTTAC	CTGGTGTCGT	GCCATTGCTA	3060
CAGCAAGCCC	AAGCACGTCA	CATTCCCTGC	GCCGTGGCTT	CGGCTTCAAA	AAACGCGCCC	3120
CTGATTTTGG	AAAAATTAGG	CGTGCGCGCG	TAcTTCCGCA	CAATTGTCTGA	TCCCGACTCG	3180
TTAAGTAAAG	GCAAACCTGA	TCCTGAAATC	TTTTTAGCCG	CCGCTGACAG	TATTGGCGTG	3240
cTACCGCAAA	ACGCCATTGG	CTTTGAAGAT	GCACAATCAG	GCATTGACGG	CTTGAAAGCC	3300
GCGGGCATCT	ATGCCGTCGG	CTTGTCGGCC	AGCCAACCCT	TAATCGGAGC	CGACATGCAA	3360
GTTTCTGAAA	TGACTGAACT	AAGCGTTGAT	GCACTGCTCA	ATCGCTAAAG	AATGACAAGA	3420
ACCTTTAGCA	GCAGGCACCT	CTTTTCATCT	ACATTGTATC	TATGCTAGAA	TGTAGATGAA	3480
AAGAGAGAGA	TGCGTCAACA	TACCTCTCTA	GTGTAGAGCC	GTTTGAGACG	GTGACCAATT	3540
TAGTTATAAA	AAATAACCGT	ACCGGTCAAA	GTAGACGGTT	ATTTTTTATT	GTCATTACAC	3600
TCCATTATTG	GATTAAAAAG	GATTGAGTTG	AACGTGGTTC	TGTGGTAGAC	TTTTATTAGA	3660
AACAAGTAGA	TGTTTCTTGA	TGATCAGGCA	TCATAATAAC	TCAATACTGA	AATCGGAAGT	3720
AATAGCCGAT	AGTTGGCGAG	CAACAGAAAA	ACTCGATATA	GAATTGGACG	TAGAGCCAAT	3780
AGTTGGTGAG	CAACAGAAAA	ACCTGAATTA	TATTTGAAGT	TTAAAGAAGG	CAAGCTAAGC	3840
TTGGCTTCTT	TATTTATAGG	AGAACTAAT	GATTGAGAAA	AAAGATTTTG	ACACAATTAG	3900
AAAGGCGTTC	TTAATTTATG	ACACCAACTT	CAATAATCAT	ATGTTCACTT	ATACATATCA	3960
ACATCGGAAA	ACAAAGAAAT	TGTTAGAAAT	GAACGTTCAA	TTTGAAGCTG	GGAATTTTAT	4020
GCATTTATGC	GGAGTTGACT	ATTATCAAAT	AGATAAACAA	gGGAGCTGTA	AATTTTCGTTT	4080
TAAGGCTACT	CAATTTTACA	GAGCTTTAAA	AAATAATAAA	GTTTCACTGC	GCGGGATTAA	4140
ACCTAAAGAT	GATGGAACCA	CAGGCCAAAA	ATTACAAGTA	ATTCCTCTAT	TAGAAATGCT	4200
TATTTCTCCA	GGTGACGAA	TATGCGATGG	AGGAAAATTT	TATAATTTAC	AGTATGAAAA	4260
AGCAATTCGT	TCAGGCAAAA	TGATTGTGGC	GTTAACATGT	AAAGAGAATA	ACAAAAAGTA	4320
CGTTCCACAA	TCTTTATTAT	CGTTAATAAA	TCAGCCACGA	AAGAGCCAAT	CGAAATCTTT	4380
AACTGAATCC	CATGAAGTAA	TTAAGATAAG	TAAGTCTGAG	CTGAATAGCA	CCTCAGTAAT	4440
TGAAGTTTAC	GATAAATTTT	AACAAAAAAG	CGGGCTAAAC	AGACTATAGG	CAAAGAGAAA	4500
AAGCCCAGAG	GTTTTTCTCC	TCTTTAACAG	CTTTTAGGAt	TTCACGGATT	AACGCTAAAT	4560

TTATTCGTAC TTTTCGGGAA AATGCGCTAT AATGAAAATG AAAAGAGAGA TATGCTTCAA	4620
CATACCTCTC TGATGTAGAG CCGTTTAAGA CGGTGACTTT TTGAATTATT TAAAAATAAC	4680
CGTACTCGGT CAAAGTTGAC GGTATTTTT TATTGTCATT TTTAACAGCT TTTTAGTTTC	4740
CAAGGACCAA TACTTTACTA TTAGCAACGA AGTTCACGGT CCTTTGTGCT TGTCTCCACG	4800
TTCATTTCTT TTGCTTATTG TTTAAATCCT ACTAACGGAA AAATGAAATC AACGAATTTT	4860
ATCCCCGCAT AAAACATAG GGCAACATC AAAATCTTAA TTAAGATTTT GACAATTGGT	4920
CGTTCAGTTA TCTTTT <sub>1</sub> CAT TTGTATACTC CTTTCCaATA TTTTAAAGC TCCAACCTTT	4980
TTAATCTTTT TGACATCTAT ATTTTATCTA ATTATAATAA AACTCACAAT AGAGGAAGCA	5040
TCATTTTTTC ATTTCTTTT TATTCCTTGA TTAGTAGTTC AATTTGTAAA CATATCATTC	5100
TGACAAAATA GAAAAGGAGG TTGTTATAGT ATCGAAACAA AATTATCTAC AAAGAAAAAA	5160
ACACTTCTTT ATCTCCTGTT CACTTTTTTG ACAGCCCTAC CCAGCATAGA AGCATTAAAC	5220
GGAATGAGCT TAATAGCTAT AAAAAATAAC GAAAAGCCT TGATTATCTT TGATAGTGTA	5280
CTATCTTTAA TTATTATTAT CGTGCTGTCA GTTTTCTATA AAAAAATAGT AAAGAAAAAA	5340
AAGCAAGCTA TCTTTAAAAG TCTATTACTA GCTATTGGCA TGTCTTATT TTCTATTATT	5400
CTAAATTTAC TTTTCTCACG CATTCCTACT CCAGAAAATC AACATTATAT TAATGAAGTC	5460
CGAAATGTAG CACCTGTTTT AATGGCCCTT CATACTAGAC TCTTTTCACC TATAATAGAA	5520
GAACTCCTGA CTCGTGGCAT TTTTATGAAC TTATTTTTTA TAAAAACaC CTCTCGAAGT	5580
GTCTTATGTA AAATATTTTT TTCTGGTCTT TTTTTTGGT CTTAGCCATG GACTATTACC	5640
ATCAATTTCC ATGCTTTTTT ATTGTGGAAT GGGCTGGGT TTAGCCTCTA CCTATTATTT	5700
CTGTAATAAC AATCTCACGT ATCCTCTTGC TATTCATTTG CTATTAAATA ATTTATAATA	5760
CACAAAAaC CCGCCACAT CAGAAAATG ATGTGGGCGG GTTTTATAAT TATTTATTTT	5820
TkGTTTCAAA TAACGGACTA ACTGGTTTGT TTCGTGGAT ACGTTTAATC GCATCTCCGA	5880
TTAATTCACC AACGCTGACT TCGTCAATTT TaACAATTTT ACGATCATCA GATAGGTAAA	5940
TAGAGTCTGT TACAAC <sub>2</sub> TAAG CGTTCAATTG CTGAATCTTC AATACGTTGT AAGGCTGGTC	6000
CTGATAAAAC GGGATGTGTA CAAGAAGCGT AAACGTCTTT TGCTCCGGCT TCTTCAATG	6060
CGTTTGCAGC AAGAGAAATT GTCCCCGTG TGTCATCAT GTCATCAATT AAGACACATG	6120
TTTTGCCTTC CACATGACCA ATGATATTCA TTA <sub>2</sub> CTTCTGC TACGTTGCT TTTGGACGAC	6180
GTTTGTCAAT GATCGCAATT GGTGCTTTTA AGAATTCTGC CAATTTACGC GCACGCGTTA	6240
CACCACCATG GTCAGGGGAA ACAACAACGA CGTCATCGCC TTGAATGCCG TGTTGATGA	6300
AGTAATCGGC AATTAATGGT GCACCCATCA AATGATCCAC TGGAATATCA AAGAATCCTT	6360
GAATTTGGAC CGCATGTAAG TCCAATGTCA ACATTCTTGT TGCACCAGCT TTTTCAATCA	6420
TGTTTGCAAC AAGTTTAGCA GTGATTGGTT CCCGGGACG TGCTTTACGG TCTTGACGTG	6480
CATACCCATA ATAAGGCATA ACCACGTAA TTGTTT <sub>2</sub> TAGC ACTCGCACGT TTTAACGCAT	6540



CAATCATGAT TAATAGTTCC ATTAAGTTAT CGTTAACAGG GCTACTTGTT GATTGAATGA	6600
CGTAAAGATG TGCACCACGA ATACTTTCTT CAATATTCAC TTGAATTTTCG CCATCACTAA	6660
ATTGTGTTAC AGAGCACTTT CCTAGCTCCA CACCCACAGC ATCCGCAATT TTTTCTGCTA	6720
AGGGACGATT AGAGTTCAAA GAAAAGATTT TCAATCTTGG ATCAAAATAA TGTTTCGACA	6780
TTAGAACCTC CACTTTCTAT TTCAGCCAAT TCGTTAAAC TGaACAATAC TThCCTTCTA	6840
AAATAATAGT AGTCCTAGCC mAATAATCa aGAGaTTTTT tGtATTTTCAT ACmAaGATAT	6900
TCmTcAAAAG tCGcTTTCTT CTAActATTT TATCGTATTT GTCTAAAAAA TGAGGATTAT	6960
TTGAATAAAG AAAGGTAATC GCCATAGCCT TCTTTTTCTA AATCTGCCAC TGGCACAAT	7020
CTCAAAGCTG CCGCATTAAAT GCAATAACGT AAGCCACCTT CTTGGAGTGG GCCATCAGTG	7080
AAAACATGCC CTAAATGTGA ATCCGCTTCT TGA CTGCGAA CTTCTACTCG GTGCATGCCG	7140
TGACTAAAT CAGCTTTTTT TTTGACGCCA CGTTTTTCAA TTGGTTTGGT AAAGGATGGC	7200
CAGCCGCAAC CAGCATCGTA TTTGTCCAGG GAGCTAAACA ACGGTTCCGCC ACTAACAATG	7260
TCTACATAGA TTCCGTCTTG GTAAAAGTCA TCATATTCTC CTGAAAAGG GCGTTCTGTT	7320
GCGTTTTCTT GTGTGACGGC ATATTGAAGA TCCGTCAAGG TTTGTTTTAA TTCTTCTCT	7380
GTTGGTTTTG TCATTCTGTC CACTCCTTTT CTTCTGTTCA ATTTATTGGG TTCAATGATA	7440
CCACGTTTGC CAAACAGGCA CCTaATAGTT TGCATAGTAA AAGACAGTCC TTTCTCTTTT	7500
TGACGAATGA AAAGGACTGT CTTTTATTAG TTGAATTCTT TTAACAGTCG AGCTACTTTG	7560
GCAATTGGTA AGCCCATAAAT GCGTAATAA TCGCCAGCGA TTGCTTCAAT TAGTAATGCT	7620
CCTTGTCCTT GGATACCATA AGCGCCTGCT TTGTCCGCAT ATTCTGCGGT GTCTAAGTAA	7680
GCATGGATTT CCGTGTCTGT CAGTGGATAA AACGTACGG TTGAATGAAC TGTCGCGGA <sub>g</sub>	7740
cGCTcTTTTT CCCCTTGTTT CAACGTCACA CTTGTATACA CGTCATGTGT TTTGCCACTT	7800
AGGAGGCGCA ACATGCGATA ACCGTCTTCC CGAGAAGTGG GTTTGCCGAG GATCTCTCCC	7860
GCCAATGCCA CAATTGTATC GCAGCCAATG ACTAGTGCTT CGGGGGATTG CTCTGCTATA	7920
GCCGCTGCTT TTTGGGCAGC CATTTGCGCC ACATATTCTG CTGGTAA <sub>g</sub> CC GTCTTTGCCA	7980
ACAGTTTCAT CGATGTCGGC TGGTGCAATC GTAAAAGTTG GAACCACACG CTTTAATAAC	8040
TCCTGACGGC GAGGAGATTG GGATGCTAAA ATAATTTGCA TAAATTGATG ACGCCTCGTT	8100
TCTTTTTAAT GGGGGTCTTG AATGCGTTTA AACATTGCTT CCATATCTGA ATTCGTAAAA	8160
TGAATCAGCA CCGGACGTCC ATGCGGACAA TTAAACGGAT TTTGCACAA GGCTAAATCT	8220
TTTAATAGCA CGCGTGCTTG TTGCTCATTG AGGTAATGAT TGGCTTTGAT GGATCGCTTA	8280
CAGCTCATCA TAATGGCAGT TGCTTCACGG AATTTTTTCA CGCTGACCGA GCCAGTTGTT	8340
AGCAACATGT CAATCATCTC GCGAACAATG CTTTCCTCTT CGCCTGCAGG ATACCAAGTT	8400
GGGTGCGCCC GCACAATGAA ACTATTTTGT CCAAAGTCTT CCAAATGGAT GCCAACTTCC	8460
GCAAGGGTTT CTTTCTGCTC TTTAATTTTC AAGGCATCGC TATTTGGATA ATCAATAACG	8520

ATTGGTACCA	ATAGTTCTTG	TAAATCATCA	CTGACTTCGC	CAATTTTTTC	ACGGAAGTAC	8580
TCGTATTTAA	TTCGTTCTTG	GGCAGCGTGT	TGGTCGATAA	TAAACAGGCC	ATCTTTGCTT	8640
TGTGCGAATA	AATAAGTTCC	GTGCATTTGA	CCGAAGTATT	CTAGCTCAGG	AAAACGCTCT	8700
GTCGGTCGTT	CTTCTTCCAA	CTTGTTTCAGG	GCTTGTTTCA	GTTCTCTTTG	ACTCGCAGCG	8760
CTATTGGCAT	CGAATTCTGG	GTGAAGTTTC	ATTTCTTCTT	GGAAAACCTC	GTCAGCAGCC	8820
TCTGATGAAC	GCAGAGAAGG	TTCTGAATCA	GGCACTTCAG	TTATTTTCGT	TTCTTCGGCT	8880
ATTACCTGTT	GCTCATTACT	AGCTGTTTCC	CAATGATTTG	CGGAATCAGC	AGCTGCCGAA	8940
GAGGTTTCGT	AAGCTGACGT	TTCTTCCGCT	ACACTCGTTG	ATTCTGGCAC	TGGCGGCGCA	9000
AATGGATCGT	CCATTGGGTC	CACTGCTCTT	TCTGTCTGGT	TGACAGCTTC	AGGGGCGCTT	9060
GAGGCTGGTG	TTTCGTCTGT	CGCATGTTCC	GCATAGAAGC	GGCCTGTTGC	GGGATTATAG	9120
CTTAAGCTGC	CCGGTCGTTG	TTTTACCGCA	GgTTGTTCAG	TCGATAAAGG	AATTTCCATT	9180
TGTTCCGGCTT	TAGGTTCTGG	CGGTAGTTTC	TTTTTAAAC	GCAAATTGTC	TGCCGCATTG	9240
GGAATCAACT	GTTCTTGGCT	CAACACTTCT	TGAATCGCCT	GTTCAATTAA	TGCCATTAAA	9300
TCTTTTCTT	TACTTAAACG	AACTTCTTGC	TTAGTAGGAT	GGACATTGAC	ATCGACTAAC	9360
AGCGGATCCA	TTTCAATTTT	TAACACAGCA	ATTGGAAAAC	GCCCCACCAT	TAGTTTCGAA	9420
CCATAGCCGG	CTACAATTGC	TTTATTTAAC	GCGAAATTTT	TAATATAGCG	ACCATTGATA	9480
ATCGTTGATA	AATAGTTGCG	GCTGGCCCGT	GTGACTTCTG	GCAATGAGAC	ATACCCTGTC	9540
AAAGTAAAT	CAAGGTCTTT	TCCTTCAATT	TTCAACATTT	TCTTTGCGGT	GCTAATGCCA	9600
TAAATCCCCG	CAATCGTTTG	TTTTAAATCA	CCATTGCCTG	TTGTACTCAT	CATTTTGTGG	9660
CCATCATGAA	CTAAGCGAAA	AGCAACTTTA	GGGTGGCTTA	ACGCCAGACG	ATTGACAATG	9720
TCACCGATAT	TGGCCAATTC	CGTTTGGATC	GTTTTGACGT	ATTTCAAACG	AGCGGGCGTA	9780
TTATAAAATA	AATTAGAAAC	GGTCATTTTT	GTGCCTTTAC	GGAGGGCTGC	TGGTCGGTTT	9840
TCTTCGACCT	TGCCCCCTTT	TAGAATCACA	TAGCTGCCTT	CTTCTTCCTC	AGCTGTGGCA	9900
GTTTCGACAA	TCATTTCAGA	AACTGAGGCA	ATACTTGGtA	AnGcTTCGCC	TCGGAACCCC	9960
AAGCTGCGAA	TTCGGAACAA	ATCATCTCTT	GTATGAATTT	TGCTTGTTGGC	ATGGCGCTTG	10020
AAGGCATTCA	GTACATCTTC	TTTGGCAATT	CCTTCACCAT	TATCAATAAT	TTGAATGGTT	10080
TTTAAGCCAG	CTTCTTCAAT	AAAAATATCA	ATTTGAGTGC	TTCTGCATC	CAGTGCATTT	10140
TCCACCAATT	CTTTGACGAC	AGAAGCGGGA	CGTTCAACAA	CTTCACCAGC	AGCAATCTGG	10200
TTGGCTAGTT	GTTCCGATAG	TTCTTGAATT	TTCCCCATTT	CACGACCTCC	TAACCGGTTT	10260
GGTTCCTAAT	TAGATACGTT	TTTGCAATTG	ATGAAGCATG	TTTAACGCAT	CTAGTGGTGT	10320
CATTTCTAAT	AGATTCATTT	TCTTCAATGT	GTCAATGACA	CTTAATTCTT	CTGTTGAAAC	10380
TTCTTTAAAT	AATGAAAGTT	GCTCGGTTTC	TTCATGTACT	TCTGAAACTT	CTTCATGATG	10440
AACTGAAATT	GGAATCGTTG	TTTCTTCCGC	TTCTAAAGCG	GATAAAATGG	TGGCTGCTCG	10500

TTCTAAAAGC GGGCTTGGTA AACCAGCGAT TTTAGCAACA TGTATCCCAT AGCTCTTATC	10560
AGCAGGTCCT TCCATCATT TATGAAGAAA GACCACTTCG CCATCTTTTT CGACAGCGC	10620
CACGTGAATA TTCTTCAAGC CTTTAAAGGT TTCATCTAGA ACAGTTAGTT CATGATAATG	10680
TGTAGAGAAT AGCGTTTTGG CTTGTACTTC ACGATGGATA TATTCAATGA TGGCTTGCG	10740
TAGCGCCATT CCATCATAGG TTGCCGTGCC GCGTCCTAAT TCATCAAATA AAATTAGGCT	10800
GTTAGGCGTG GCGTGCCGCA GTGCTTGGTT GGCTTCCATC ATTTCTACCA TAAAGGTACT	10860
TTGCCCTGCA ATTAGATCAT CAGAGGCCCC AATCCGTGTA AAAATTTGAT CGAAAATTGG	10920
CATTTCAAGCA CTTTCCGCTG GTACAAAACA GCCAATTGTC GCCATGACCA CCGTCAAAGC	10980
CAATTGACGC ATGTAGGTAC TTTTCCCTGA CATGTTTGGT CCAGTAATCA AGAGAATATC	11040
TGTTTCTGGG TTCATGCGGA TACTGTTTGG AATGTATTCT TGATGCCCTA AAACCTTTTC	11100
CACGACTGGG TGTCGTCCTT CGACAATCGC CAAATTTTTC GTGTTACTTC TTAACGTTGG	11160
TCGAACGTAT TGATACCGTT CACTAATTGT TGCAAAAGAT TGTA AACAT CAACCGCACT	11220
AATGGTTTTA GCCAGTGTTT GCAACCGATC AATGTTGCTT TCACTTGTT CTCTCACGGC	11280
CAAGAACAAT TGATATTCTA GTTCAACTGA TTTTCTTCG GCTTCTAAGA TTAAACGTTT	11340
TAATCTTTT AGCTCTGGCG TAATAAGCG CCGCATTG GCTAATGTT GCTTACGCTC	11400
GTATTTTCCT TCTTCTAAAT TCGCTAAATT TGACTTAGTG ATTTCAATGA AATAACCAA	11460
GACCCGATTG TAGCCAATCT TCAAGTTTTT AATCCCTGTT TCTTGCGTT CTTTCGCTTC	11520
TAACTCCGCT AACCATTGCT TACCATTGCG CATGGCATCA CGGTATTCAT CTAATTGGTC	11580
GTTATAGCCG TCTTTAATAA CCTTTCCTTC AGTAATTTGT AAAGGGGCTT CTTTATTAA	11640
CGCCGTCGCA ATCAATGCCA CTAAATCTTC AACAGGGTTT AAGTCGACTA ACAAATCATC	11700
CCATCCCCCT TGATTGATCC CGACAATTAA CTGTGCAATC GTTGGTACTT GTTCTAAAGA	11760
GGTCCGTAAT TGAATTAAAT CACGCCCAT CACACTGCCG AATGCCACGC GTCCTGCTAA	11820
CCGCTCTAAA TCATAACGT TCGTCAATGC TGCTTGCAA TCCAAGCGTT CAAAATAGGC	11880
ATTTAATAAG GATTGCACCA TTTCTTGACG CGCCTTAATT TGACGTTCTT GAATCAGCGG	11940
ACGATCCAAC CATTGTTTTA ATAAACGGCC GCCCATAGCT GTTTTCGTT CATCTAATAG	12000
CCACAATAGA GTCCCTTTTT TCTGTCCTGT TCGAATCGAC TGAATCAATT CTAAATTGAA	12060
TTTAGAATAA TGATCCATTT TCAAGAAATG ATCTGGCTGA TATTCCACAG CTTTCTGAAT	12120
ATGGGACAAG CTCCGCTTTT GGGTAACGGA TAAATAAGAT AATAATTGTC CTGTAATTTT	12180
AATTTCTAAT GGATTAATCA GTTCATTGGT TAAAAACTA AACTCGGCAT TTTCTTCCAC	12240
CGTTTCTTGT GTCGAAAAGA TGATATTTAA ACGTAAACTT AAGTTCTCTT TTAACGATTC	12300
AGGGATCCCT GAACCTAGGA CCATTTCTTT TGTTTGCAAC GCAGAAGCTT CGTTCAAAC	12360
GCCCTCTTCA TCAGCCAAGA CTGCTGTTTT TAATTCGCC TACTCAAAT CTACATAGGC	12420
AAAACCGAAC TGATTACCAT CTGTTAGCAC AGCGGTTAAG TAATTATTGT CTTTGGCTTC	12480

CAAGCCTTTA	CTATTCATCA	CCGTTCCCTGG	CGTTACCAAC	TGGACCACTT	CTCGTTTGAC	12540
CATGCCCTTG	GTAGTCTTTG	GATCTTCCAC	TTGCTCACAA	ATCGCTACTT	TATAGCCTTG	12600
CTCAATCAAC	GTATCAATAT	AGCCTTGCGC	GGCATGATAA	GGAACGCCAC	ACATAGGAAT	12660
CGGATCATCC	GCATTGCGGT	TACGACTGGT	TAACGTCAAT	TCTAAGATTT	GCGCTGCATT	12720
GATAGCATCT	TCATAAAATA	ATTCATAAAA	ATCACCTAAG	CGATAAAATA	AAAACGCATC	12780
CTTGATTGG	GCCTTAATGG	ATAAATATTG	TTCCATCATG	GGTGTATTTT	TTGTCTTTTG	12840
TGGCATTTAA	TTCCCTTCCT	TCTCTAGCTC	TTCGTTACAG	CGGTATTGAA	TGAGCGCCGT	12900
GACATGTTGG	ACTAAGTCAT	TGGCTTCGAT	CAGTTGTTCC	CGGTAGGCTA	CAACGAGTGG	12960
GTGCTCATCA	AATTCTTTTG	TTTTCGCATC	GGCTCGTTGA	ATGGCTTCTT	TTTCAGCAGT	13020
TGGTTTGCCA	TAGTGAGCAA	ATTGAACCGC	ATCTTTTTGC	GCAGCTTTAA	TTTCTTCCAC	13080
TAATTCTGTC	AGCTTTTTGT	TTTTCTTTAC	CTTTTCTTCA	ATCGCTTGGT	AGCGCTGAAT	13140
GACTTCATCT	TGTCCAATTA	AATGTAACAA	TTTTTCTAAT	TCTTTATTAC	TTTGTGCATC	13200
TAAATTCAAC	GGCTCCATTT	TTTTCACCTT	CATCTTTCAC	CGGCACTTTT	TCTATAAGAG	13260
CAGAAAAAGT	TAACCTCTTA	ATTATCCGAG	ATTAACTTTT	CTTCTATTAT	TCTTGAAACG	13320
GGCGATCTTC	GCTAATCTGA	ATCGGCTCAA	TCATTTTCGC	TTGGCCCGTT	TGGTCATCAA	13380
TCTCTAAAAT	ACAGGCAGAT	AAAATTGTCC	GCCCTGTTTC	GACTIONTCA	AAGCGCTTCG	13440
GCAATGCTGT	CATAAACTTC	TCAATAATCG	GTTACGCTT	CATTCTTAAA	ATCCCATCAT	13500
AAGGACCGGT	CATACCGACA	TCACTTAAGA	AGGCTGTCCC	TTTCGGTAAA	ATCCGAGCGT	13560
CATTGTGTTG	AACATGCGTA	TGTGTTCCAA	CAACGGCTGA	AACTTGGCCA	TCTAAAAACC	13620
AGCTCATGGC	TTGTTTTTCA	CTCGTCGTTT	CGCCATGAAA	ATCCACAAAA	ATGAGCGGTG	13680
TCCGTTTGCG	AGCGATAGCT	ACTAACTCGG	CGGCTTTCTT	GAAAGGATCT	TCCAATGGCA	13740
CCATAAATGT	GCGGGCCTGT	AAATTAATCA	CCGCTAGTTC	CAATTGATTG	ACTTTCACAA	13800
AAACCATTC	TTGACCAGGC	ACATCTTCTG	GAAAGTTCGC	TGGACGAACC	ATTTTCTTCG	13860
CATCATCAAT	GAATTCAAAG	ATATCCTTGT	TGTCCACCGT	ATGGTTACCC	ATCGTAACAA	13920
CATCGACACC	ATCTTGCAAA	AATTTCTTAT	AAATCTTGCC	AGTAATTCCC	CGACCAGATG	13980
CCGCATTTTC	ACCATTTGGCA	ATGGTCACTT	GTGGACGATA	TTTTTTCTTT	AATTTAGGGA	14040
CATACGTTGC	TAAGGCTTCT	CTTCCTAAAG	AACCGACAAC	ATCTCCTATA	AATAATATAC	14100
GCAAAAAGGT	TCCTCTTCTC	TTTTTAACAT	CTCTTGTTTA	TTATAGTTGT	TTTCGCTTTG	14160
AAAAGAAAAGT	TTTGTTTTTA	AAAGATTAAA	AAAAGTGTCT	CAACGAACAA	TCCTTGCAAT	14220
TGTTGTTTGA	GACACTACAC	GCGATACTTA	TTGCTTTTTT	ATTTTATTTC	GCATAGTCTA	14280
CGGCACGTGT	TTACCGGATA	ACCGTTACTT	TGATATGTCC	TGGATAATCC	AAGTCATCTT	14340
CAATTTTCTT	CCGAATGTCT	CGAACTAAAC	GAACAGATTG	TAAATCAGAA	ATTTCTTCTG	14400
GTTTCACCAT	GACACGAACT	TCACGTCCTG	CTTGAACAGC	AAAACCTGAA	TCAACGCCTT	14460

CAAAACTATT	AGAAATATTT	TCTAAGTTTT	CTAAGCGGCG	AATGTAGTTT	TCTAACGATT	14520
CcTACGTGCA	CCTGGACGGG	CTGCAGATAA	CGCATCTGCG	GCTGCCACTA	AAACAGAAAT	14580
AACTGAGGTT	GCTTCCACAT	CACCATGATG	CGAGGCAATG	GCATTAATAA	CTACTGAATT	14640
TTCTTTATAT	TTCGCAGCTA	ATTCTGCGcg	AATCtCAACG	TGAGAACCTT	CAATTTTCATG	14700
ATCCAATGCT	TTCCCAATAT	CATGTAGTAA	ACCTGCACGT	TTGGCTAATT	GAACATCTTC	14760
GCCTAATtCG	GCGGCTAAAA	CGCCAGCTAG	TTTTGCTACT	TCGATTGAAT	GGTTCAAGAC	14820
ATTTTGACCA	TAACTTGTTT	TGAAACGTAA	ACGGCCTAAA	ATTTTAATTA	AATCAGGATG	14880
TAACGTATGA	GCACCTACTT	CAAAGGCTGC	TTGTTCTCCG	TATTCACGAA	TCCGTTTCATC	14940
CATTTCTTTA	CGTGATTTTT	CTACCATTTT	TTCGATACGA	GCGGGATGAA	TCCGACCATC	15000
TTGGATTAAT	TTTTCCAAGG	TCATTCGAGC	GATTTCTCGT	CTGATTGGAT	CAAATCCGGA	15060
TAACACAACC	GCTTCTGGCG	TATCATCAAT	GATTAAGTCA	ATCCCTGTTA	ATGTTTCCAA	15120
TGTACGAATA	TTACGTCCCT	CACGACCAAT	AATGCGACCT	TTCATCTCAT	CATTTGGTAA	15180
AGTGACCACG	GAAACAGTTG	TTTCAGAAAC	AGAATCGGCG	GCGCAACGTT	GAATTGCTAG	15240
GGAAAGCAAG	TtcTTCGCTT	TGCGATCAGA	TTCTTCTTTG	gCACGTTGTT	CTGACTCCTT	15300
GACCATCAAG	GTTAGTTCAT	GATTCAACTC	TTCTTCTGTT	GACTTCATAA	TAATCGATTT	15360
TGCTTCTTCT	TTTGAAAGGG	AAGCAATTCT	TTCTAATTCT	TGATGTTGCT	GATTAATTAA	15420
GTTTTCTACT	TCTTTTTCTC	GCTCATCAAT	TAATTGCTGT	CTAGCACCAA	GTTTCTCTTC	15480
TTTTTCTTCC	AGTGAGCCTT	CACGTTTTTC	mAGrGAGTCA	TCTTTACGAT	CCAACGTTTG	15540
TTCCCGTTGA	ATTAAACGAT	TTTCTTGAGA	TTTAAGTTCT	AATCTGCTTT	CTTTCAACTC	15600
ACTTTcAATT	TCTGAACGAT	ACTTCTGGTT	CTCTTCCTTA	GCTTCTAGCA	ATGCTTCTTT	15660
TTTCAGTGTT	TCTGATTCTT	TcCGAGCGTT	CTCGAGAATA	cTCGACGCTG	AAATTTTTGC	15720
ACCATTTATT	GCTTTATCGT	GACGTGACTT	TGCAATAACA	AATCCTAAAC	CAAGACCGAC	15780
AATCAAACCG	ACGATAGCGA	GGAGAATATT	TAAAACCATA	TTCCACCTC	CATACTATCT	15840
TTTTGTAAAT	TCATATCTTG	TAGTTTTTTG	ATAAACTAAA	ATCAATTATC	AATATGCCTA	15900
CTTGCAATATT	TGTGCATACA	CACAATTTTA	TTCTAAtGTT	TGTCGTGAAA	GGTGTCAACT	15960
TAAAAAAGCC	CTTTCTTTTG	GAAAGGGCrG	CAAATATTAG	ACAACGTCAC	ATTTTTAGGT	16020
TTCCGCCGTC	ACGAATTGTT	TAAATTTTAG	AACAAACGCC	GCATAGTCTG	CTGTTTGACT	16080
AAGCTGCTTG	ATGTGGATGG	GTTCCGACAT	GGACAACAGT	CTATTTAACA	ATTTAGTTCT	16140
AAACAAAAAA	AGACAACACC	TGTCCTGTTT	AGTAAGTGTT	GTCTTTCTTC	TTAACGAGTC	16200
GTATTCACTC	GTTATCTATT	CATCTAAGGG	TAATTCTTCT	TGACCTTCTG	CTTCTTCTGT	16260
AATGGTTGAA	CCGTCACCGA	TGCCGTATGC	ATCACGGACC	AATTTAGACA	CTTCTGCCAT	16320
CATTTCAGGA	TGATCCGCCA	TGTATTGTTT	GGCATTTTCA	CGGCCTTGGC	CAATTCGTTC	16380
TTCTTTGTAG	CCATACCAAG	CACCACTTTT	ACTGATAAGG	TCTTTTTCCA	CAGCCATATC	16440

TAGTAATTCG CCTTCTTGGG AAATACCTTG GCCATACATG ATATCAACTT CTGCCACTTT	16500
AAATGGTGGC GGCACCTTTGt TTTTAACGAC TTTaATTTTT GTGCGGTAC CGACGATGtC	16560
TGTTCTTGT TTTAACTGTT CTGCACGACG GACTTCTAGA CGAACCGTTG CGTAGAATTT	16620
CAATGCACGT CCACCAGGAG TTGTTTCAGG ATTTCCGAAC ATCACGCCAA CTTTTTCACG	16680
AATTTGGTTA ATGAAAATAG CAATTGTTTT TGTCTTATTA ATTGAGCCTG ATAATTACG	16740
TAGTGCTTGA GACATTAGTC GAGCTTGTA GCGGACATGG CTCGCTCCCA TCTCACCATC	16800
AATCTCTGCA CGAGGAACTA ACGCAGCAAC CGAGTCGATG ACAACGATGT CAATCGCACC	16860
ACTTGAACT AAGGCATCGG CAATCTCTAA GCCTTGCTCG CCCGTATCTG GTTGAGATAA	16920
AAGTAATTC TCGATGTTAA CGCCTAGTTT CTCCGCATAT TGAGGATCCA ATGCATGCTC	16980
AGCATCGATA AAAGCGGCTG TTCCACCATT TCGTTGCACT TCTGCAATTG CGTGTAAGA	17040
AACAGTTGTT TTACCTGAAC TCTCAGGACC ATATACTTCA ATAATTC	17087

## (2) INFORMATION FOR SEQ ID NO: 48:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11764 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

AAATGAATAT TTATTCTGCT TTATTnTCAA TACTTTGTTT TCATAGTTGC TTTAGCTGTT	60
CCTTTTTTGA ATAAGATTAA AATGAATAAA TAACAAGCAA AATATAAGAT TTTTTTGTTT	120
AACTCTTTTT GTTCAATAG GAAGTGCTTA GTGTTTTTTC AGAATTGAAA AATGAATACG	180
ACAAAAAATT TATCTATATA ATAGAAGGAA AAATAAAAAT AGTGAATGAT AAAAAAATT	240
CAAAGAAAAC ACCCGTTCTT TTTAGGCCTG AGTTTTTGAA AAAAACATTC TCATGTCTAG	300
TATCTTTAAC ATTTTGTTAC TTTTGGACTG AAACGAATGT AAAAAACCG ACATTTAAGG	360
GCTAATCAAG GGTAGAAAA CGGTTTCTCG AAAATAAAAT CCTGTTTTTT TAACTTTTTT	420
TCGCAAAATA TAAACAGGAT AATATCAAGC TTCCAGCAGC TTTTATTAC AAAATGTAAC	480
AAAACCGACA TTCTAGTAAT AAATTGTCAC GAGATATTCA CtACGGCCGG CCATTGATG	540
ATAAAGTATG TGAAGTCGTT TCGATAAAGC GAAGCAGACG TGGTTGTTGC CGAGTGCTAA	600
GAGCAACAAT CAAATATTTT TAATTACAAA GGAGAAACAT ACATGAAATC AATCAAAACG	660
ATTTTACTTG GAACAACTTT GGCCGCAGtT TAGGATTATT CTTAGGAACA GACGCAAACG	720
CAGAAAGCTT ATATACAGTT AAAGCAGGAG ATACTTTATC AACAATTTCT CATCAATTTG	780
CAGGAGACAA TAGCTTAATT CAAAAAATTG CTTCTGATAA CAAATTGCCA AACCTTGATT	840
TAATTTTTGA AGGAGAACAA TTAGTTATTC GTTCAGAAAA AGAaTTGCT AATACTCCAG	900
CACCAGCTGT AGAAGTTGCA CCAGTTCAAC AAGTAGTTGA ACAACCTGTT GCACAACCAG	960
TACAACAAGA AGTACAACAA CCTGTTGCTC AAGAAGTAGC GCAACCAGCA GCACCTGCTG	1020

CAAGCAGTGA	TGCAAAAGAG	TGGATCGCAC	AACGTGAATC	TAGTGGTTCT	TACGATGCAA	1080
CAAACGGTCA	GTATATCGGT	CGTTACCAAT	TATCTGCCTC	TTATTTAAAT	GGTGAATATT	1140
CACCTGCCAA	CCAAGAACGC	GTGGCAGATC	AGTACGTAGC	AGGTCGCTAT	GGTTCATGGG	1200
ATGCAGCTAA	ATCATTCTGG	TTAGCAAATG	GTTGGTACTA	AAATTAAATA	AACGArTAAT	1260
CTAATTACTT	ATTAAAAACC	ATAAACAAAA	AAACAGTAAC	CTGTAAAGAT	TACTGTTTTT	1320
TTTGTGATT	TGTTAAGGTA	CAACGCCCAA	CGCTCGTCAA	TTGGTGGTCT	GCGTTATAAA	1380
TTTTTGCTTC	CCAAACTTGA	AGGGTCTTTC	CCGAGTGATC	GGGCGTCGCA	ATCACGGTAA	1440
GGCTACCATC	ATGGACACTA	CGAAGATGAT	TTACTTGTA	GTCTATGCCG	ACAGCAAAAG	1500
TGTTTTCAGG	AACGTTTTCG	TCGCACCTA	AACTACAGGC	AGTTTCAATT	AAGACGCCGT	1560
TTAAGCCCCC	GTGGACAAGG	CCATAGGGTT	GTTTATGCAC	GTCTTTTACC	GTTAAAGTAA	1620
GCTGACATTT	CTCAGCAGAA	ACCTGTTGAA	TCGTGATGCC	TAAATACTCT	AAAAGATGAA	1680
TAGTAATCaG	TCCTTTCTAA	TAAAAGAGAG	AATAAAAGGT	TTTTaTTaTT	ATACAGGAAT	1740
GAAGAAAATT	kGAAATAGCC	AKGTTTTtC	TTAAATtCa	AGCGAmTTTT	TCyCtAAAGg	1800
CCGATTCTgk	GkATAwTrGA	AaGATGrGA	TtAATGAAGG	AGGTTTCCTA	ATGAGAAACT	1860
GGACAACAAT	TAAAGAATAT	GATGAAATTT	TATTTGAGCA	AGCAGGCAAA	GTAGCAAAAA	1920
TTACTATCAA	CCGCCCGCAT	GTTCAACAACG	CATTACACACC	AAAAACAGTC	ATGGAAATGA	1980
TTGATGCATT	CAACATTAGT	CGTGATAAAG	AAGATGTCGG	GGTTATTATT	TTAACAGGCG	2040
CAGGTGACCA	AGCATTCTGT	TCTGGTGGCG	ATCAAAAAGT	TCGAGGTAAT	GGTGGTTACG	2100
TTGGCGAAGA	CAACATTCCA	CGTTTAAATG	TGTTAGATTT	ACAACGGTTA	ATCCGTGTCA	2160
TTCCCAAACC	AGTCATCGCA	ATGGTGAAAG	GCTGGTCCAT	TGGTGGCGGC	AATGTTTTAC	2220
AATTAGTTTG	TGATTTAACA	ATTGCGGCGG	AAAATGCTAA	GTTTGGGCAA	ACAGGACCAA	2280
ATGTCGGTAG	CTTTGATGGT	GGATACGGCT	CAGGTTACTT	AGCGCGTGTT	ATCGGCCATA	2340
AAAAAGCAAA	AGAAGTTTGG	TTTATGTGTA	AACAGTATTC	TGCTCAAGAA	GCTTTAGATA	2400
TGGGCTGGAT	TAATACAGTT	GTTCCATTGG	AACAAGTAGA	AGATGTGACA	ATGGAGTGGG	2460
CAGAAGAAAT	GTTAACCAAA	AGCCCGATTG	CTTTACGGAT	GATTAAAGCC	TCTTTAAATG	2520
CCGATACAGA	TGGCTTGGCT	GGTGTGCAAC	AATTAGCTGG	TGATGCGACA	CTTCTTTACT	2580
ATACAATGGC	TGAAGCACAA	GAAGGCCGAG	ATGCATTTAA	AGAAAAACGG	ACACCAGATT	2640
TCGATCAATT	CCCTAAATTC	CCATAAACCA	ATGACTTGGT	TAAATAAACA	AGTACAAAAA	2700
CGTCCCGATC	ATCCTGCTTT	TTATTTTCAG	GATGAATCTT	GGACGTTTTT	AGAAGTGcAA	2760
CAAGAAGTCA	GCCATTGGGT	AGCCACATAT	CAACAGGTGC	TTGCACCAGA	AGAAAAACGT	2820
GTGGCTTTAT	TCAGTAAGAA	CAGCAAAGAG	TTGTATTTTT	CAATTCTAGC	CTTATGGGAA	2880
TTGGGGAAAG	AGTTATTATT	TTTAAATACG	CATTTGACAC	TTGCTGAAC	GACTTTTCAG	2940
TTAAAAGATG	CGCAAGTTAA	AACGATTATT	GGTGCCTG	AAACGCAGGC	TTTGTTAGAA	3000

GAAATTTCTT TTGTCGACGT TCAGCCAATG ATTAAAAAAC AGCATAGTCT TTCACATCAG	3060
GAGTTTCAAC AGCCAAGTGA CTTAGAGTCG GTGGCTTCTA TTATGTATAC ATCTGGAACT	3120
ACAGGACAGC CTAAAGGGGT CTTGCAACGG TTTAAAAATC ATTTGGCCAG TGCTAGAGGC	3180
ACTCAAGAAA ATATGGGGAT TACTGCTGAA GATTGTTGGT TGTGTGCAGT CCCATTATTT	3240
CACATTAGCG GATTGTGCAT TGTGTTCGA CAACTCGTCT TAGGTTGTAG TATTCGCCTT	3300
TACGATAAAT TTGATGAACA GCAAGTGACA CAAGATTTAC AAGAAGGCCG TGGAACCGTC	3360
ATTTCAAGTGG TTGCGACAAT GTTACAACAA TTATTGTCTG TCTATCCTGA AGCTGGCTAC	3420
AGCGCCAgCT TTaAAGGAAT GCTATTAGGC GGCGGACCGA TTGCTCCGGA TAAATTAGCG	3480
CAGTGTGAGG AAAAAGGGAT TCCAGTCATT CAATCTTACG GGATGACGGA AACGTGTTCTG	3540
CAAGTTGTTG CCTTGAAATT TGAAGATGCG GCACTGAAAA TCGGCTCTGC TGGACAACCG	3600
TTAAAAGATA TGCAAAATCAA AATTGTTGAT GAACTAGGAC AAGAACAGCC AGAAAAGCAA	3660
GTCGGGGAAA TTTTACTTAA GGGACCAAAC GTTGTTCAG GTTACCTTAA CCAGCGTCAA	3720
CCTGAAAAAT GGACAGCAGA TGGTTGGTTT AAAACAGGCG ACATGGGCTA TTTAGATGCA	3780
CAAAGTTACC TGTATTTGGT TAGTCGTTTA AGTGAACCTA TTATCTCTGG TGGTGAAAAAT	3840
ATTTACCCCA CTGAGGTTGA ACAGGTGTTA CAGGCGATAA CAGGAATTAA AGCAGCGGCA	3900
GTTGTAGGAG AACCATATGC GCAATGGGga GCCGTCCCAG TTGCATATGT GATTAGTGAC	3960
CAAGAAATCA CCTTAGCGCA AATTCAAGAC CAGTGTTTAC GAAAACTGGC AAAATATAAA	4020
CGACCGAAAC GCATTTATTT TTGTCATTCT TTCCCGCAA CAGCAAGTGG AAAGATTGCC	4080
AAACATCGTT TCATGACAGA AGAAAGAGAG GCGTTTTTAA TCAGATGAAA TTGCCTGATG	4140
AATTACGACA AGCCTATCAG AAGAATGCCC GTCAGTTCAG TTGGGTTTTT CCATTAGAAA	4200
CAGAACAAAC AGCCTTGCCA ATTTTGGCCG CTGGTGAAAA GGCATACCAA GGAGAACGTT	4260
TTTTTTGGCA AACGCCGAAA AAAGATTTTG CTTTGGTTGG CTTCGGATAT GAAACTATCT	4320
TAAAAGGCGC AGAAACAGAA ACGCATCAAT TGAACAACTT TTTAAAAAAA GAGGCGGCAA	4380
CTCGTTTTCA AAATATGTCT ATTTCTGGTA CGGGGGCCCT TCTGTTTGGC GGTTCGCTT	4440
TTGATACAGA ACAAACGCCT ACTGAGGCCT GGGGAGAGTT AGGGGAAGGT TGGTTTTATC	4500
TACCAAGTAT TCTATTTACT TTTTCAGGTT CACGTATTTA TGGAACGCTG AACTTTTCAG	4560
GTTCTAGTGA ACAAGAAATC GAGGAACGTT GGTGACTTT GGTGGCTCAG TTTGATCGCT	4620
TGTTAGCTAC CTGTGAGGAG TTGCCAGCTG TAGCGGAAAC AAAGATTGAC AAGGAAGAAG	4680
TTGCCGTGAC TGAATGGCTA CAAGCTGTCA ACGAAACAGT AGCAGTCTTA AGAGAAGATG	4740
GCCCTTTAAA AAAAGTGGTT TTGGCTCGAC AACTTGCTGT GAGCAGTCCT GAAAAAATTC	4800
GAAGTAACCA AGTGCTCGTT AATTTAATGG CACAACAACA AAATACGTAT TTGTTTGCTT	4860
TAGAAGCCAA AGACACAAGT TTCATTGGTG CAACACCAGA ACGACTTCTG TTGGGTACTA	4920
AAGAAACGTT TGCGACCGCT TGTATTGCAG GAACCATTAA AACGGGTCAA ACACCAGAAG	4980



AAACGAAAAC	TTAGGAGCA	CAACTTTTAC	AAGACCGCAA	AAATACCGGA	GAACACCAAA	5040
TTGTGGTGGA	ACGTTTAGCG	AAAGAGTTGG	CGAAGATGAC	GACTTCTGAA	AACAGTATTC	5100
AAGCACCCAT	TATTTTGGA	AATCGAGATG	TACAACATCT	TTATGTGCCA	ATCTCGGGTC	5160
AACGTAAACC	GGGAATTTCA	TTTTTGAAA	GTGTGATGCA	GCTACACCCA	ACGCCAGCTT	5220
TAGGTGGTGA	GCCCAAAGAG	TTGGCAGTCG	AATGGATTCG	ACAATATGAG	CCAGGGAGTC	5280
GTGGGTTATA	TGGTGCGCCA	ATCGGTGGA	TTTCAGGGAA	TGATGATAGT	GGCGAGTTCG	5340
CTGTGGCCTT	GCGTTCAGGT	GTCTTTGCTG	GTCAGCAAGG	CGTTCTTTAT	GCAGGTTGTG	5400
GTATTGTTGC	GGATTCCCAA	GCAGAGCTAG	AAAGAGAAGA	AACGAAAATA	AAATTTCAAC	5460
CGATGTTACG	AGGAATTGGA	GGTCAGGTTT	AATGAATCAT	CAAGAAACAA	TGACCGATTA	5520
TTTAATGGCC	TTTATTGAAG	GCTTAAAAA	TAGTGGGGTT	GAACAAGCAG	TCATTAGTCC	5580
AGGGTCCCGG	TCCACACCAC	TGGCTTTATT	ACTTCATCGG	GAAACCGCAA	TCCAAACTTT	5640
TGTCGATGTC	GATGAACGTT	CCGCTGCTTT	TTTTGCTTTA	GGATTAAGTA	AAGCGAGTCA	5700
AAAACCAAGT	GTCTTACTTT	GTACCTCTGG	AACAGCAGCT	GCAAACCTATT	ATCCAGCAAT	5760
TTGTGAAGCC	AATATTAGCC	ATGTGCCTTT	AGTCGTTTTA	ACAACAGATC	GGCCGCATGA	5820
GTTACGTCAA	GTTGGTGCTC	CTCAAGCCAT	GGACCAATTA	CAAATGTATC	AAAATCACGT	5880
GAAATTATTT	GTTGAAATGG	CTTTGCCAGA	AGCAACAGAA	GAGATGTTGA	ATTACGCCTA	5940
TTGGCAAGGC	GCTAAAGGAG	CAGCATTTCG	GCAACAAACG	CCTGCAGCAC	CTGTTCATTT	6000
GAATTTTCCT	TTGCGGGAAC	CATTATTACC	TGACTTAGAG	AGAAAAACAA	AGAGTTCTCA	6060
ACAAACGGCT	TTATTCGCTG	GTCAATCAAT	CCTTTCTACA	GAACAAGTGC	AACAGTTAGC	6120
GGATCAATGG	TACCAGAAAA	ACGGCGTCTT	GGTTGTCGGT	GGCAGTCATA	CAGAGGAAGA	6180
AGCGACGTTA	TTTATTCAAT	TAGCCGAAGC	TTTGCAATGG	CCGTTACTAG	CGGATCCTTT	6240
AGCAAACATT	GTGACACATG	GGCAAATAG	TGAAGTTGTT	ATTGCACACA	GTGATCTGTT	6300
TCTAAATGTG	GCGACCTTGC	CACAAGAACC	AGAAGTGGTA	GTGCGTTTTG	GTTGCTGCC	6360
AATTTCTAAA	AACATCATGT	TATGGTTGAA	ACGCTTAGCA	ACGACAGAGA	CTGCTTTTTA	6420
TTTTGTGGAC	GAAAATGGGC	AATGGCAAGA	GCAATTGAAA	AAATCACAAA	CAGTTATTCA	6480
GGCAAAGGAA	ACCACGTTTG	TAGAACAAC	ACTAACGGTT	GTCAAACCAA	CAGAAGCGAC	6540
TTGGTTAGCA	CAATGGCTCC	TGCTGGAAAA	AACGGTCTCT	GAAGTTCTTC	TAGAAACGCT	6600
GAACGCGACA	GAATTAAATG	AAACAACAGC	CAGCTTAGCT	GTTACACCAA	CAATGAAAGA	6660
AAATGGTCAG	TTATTTGTAT	CTAATAGTAT	GGCTATCCGC	TACTTAGATC	GATTTATGGA	6720
TAGTCGCCCT	TATCGGATGT	TTGGAAATCG	CGGCATCAAC	GGGATTGATG	GGATTGTTTC	6780
AACTGCTTTA	GGAATGAGTG	CCATAGCACC	GACACAGCAA	AACGTTCTAC	TGATTGGTGA	6840
TTTAGCGCTA	TATCATGATA	TGAATGGTTT	ATTTTTGGCT	AAACGTTATC	AGTTGCCGTT	6900
AACAATTGTT	TTATTAAATA	ACAATGGTGG	CGGGATTTTC	TCTTTTCTAT	CTCAACGAAC	6960

TTTACGTGAG GACGATTTCG AGCCTTTATT TGGAACACCG CTTGATTAG ATTTTTCGTT	7020
AGTTGCTGAA CTTTACGGTG CTTCTACCA AGAAGTCAAA ACAATCGCGG AgCTCAAGCA	7080
AATTTTACAA GCAGCGGCTG AAGAACCACA ATTTCAAGTG ATAGAAGTCA AAGGAAATCG	7140
CCAAGAGAAT GTTCATTTGT ACGAATCAAT TTTAGCTGAA ATTGGTAGAC GAGTAGAAAG	7200
ACAGGGGATT TCATGGAACG GTTAATTCGG GGGATGCAGT ACCATTATCA ATGGTTAACA	7260
CCTTTTGATG CAAAGCGAAC GACAGTAGTC TGTTTACACG GTTTCACAGG AACGTTAGCC	7320
ACGTTTGCAG CCGTTTTTCC CAGTCAAACA CCTTATAACG TTTTAGGAAT TGACTIONT	7380
GGGCATGGCG CGACTGCTAG TTTGGTAGCG CCAGAGCGTT ATACCATGAA ACAGGTTTGC	7440
CATGATATTG CTGAACCTAAC CGAATCGCTA AATTTGCCTT GCTTTTGTTT ATTAGGCTAT	7500
TCAATGGGCG CTCGAACAGC TTTAGGTTTT GCATTACATT ATCCACAAAA GGTACAGCAT	7560
CTTTTATTGG AAAGTGGATC ACCAGGTTTG GCCACTGCCG CAGAACGTCA GGCTCGTATC	7620
TGTCAAGATC ATCGTTTGGC GGAACGTCTT TTAGAGGAAC CGCTGGTGGA TTTTATTGAT	7680
TTTTGGCAAG AATTACCGTT ATTTCAAACG CAAAAAGCCT TGTCTGTGGC ACAGCAAATG	7740
GCCATTCGTC AGGAACGTTT GAGCCAATCG GCCTTTGGAT TGGTTAGTAG TTTGTGGTAT	7800
ATGGGCACAG GAGCGCAAGA GAGTTACTGG GAACGTCTAG CAGAATTGCA GCCGATTCTT	7860
ACCGATTTAT TGGTAGGCGG CGAAGATCAA AAATTTATTG GAATTGCGAA AAAAATGCAG	7920
GCACGTCAAC CGTTGTTGAG GCTGACAATT TTTCTGAAG CAGGACATTG TATCCATTTA	7980
GAACAACCAA CGATTTTCTA TGAAAAGGTG ACAGCGTTAT TGGAAGGAGC CATCTGAATG	8040
AACATTCAAT CAATTGAAAC CTATCAGGTT CGTTTGCCTT TAAAAACGCC GTTTGTCACG	8100
AGCTATGGTC GCCTGGAAGA AAAAGCGTTT GATCTATTTG TAATAACGGA TGAGCAAGGA	8160
AATCAAGGGT TTGGCGAATT AGTGGCTTTT GAGCAACCTG ATTATGTGCA AGAAACGTTG	8220
GTAACCGAAC GCTTCATCAT CCAACAGCAT TTAATCCGT TACTTTTAAAC AGAAGCGATT	8280
GAACAGCCCG AAGAGGTTTC AACAAATTTT GAAGAAGTTA AAGGGCATTG GATGGGAAAA	8340
GCTGCGTTAG AAACCTCCAT CTGGGATTTA TATGCCAAAC GGCAACAGAA GAGTTTAACT	8400
GAATTTTTTG GACCAACGCG GAGGAAAATT CCTGTAGGGA TTAGTTTAGG GATTCAAGAA	8460
GACTTACCTC AATTACTTAA ACAAGTTCAG TTAGCTGTAG AAAAAGGGTA TCAACGAGTG	8520
AAGCTGAAAA TTCGTCCCGG TTATGATGTG GAACCTGTTG CTTTGATTCTG TCAGCATTTT	8580
CCTAACTTGC CTTTAATGGT CGATGCAAAT TCAGCGTATA CATTGGCAGA TTTGCCTCAA	8640
TTGCAACGTT TAGATCACTA TCAATTAGCG ATGATTGAAC AGCCTTTTGC TGCAGATGAT	8700
TTTTTAGACC ACGCACAATT ACAAAGAGAA CTGAAGACAA GAATTTGTTT GGATGAAAAAT	8760
ATTCGTAGTC TCAAAGACTG CCAAGTGGCG TTAgaATTGG GCAGTTGCCG TAGTATCAAT	8820
TTGAAGATTC CTCGTGTAGG TGGGATTAC GAGGCGTTGA AGATTGCCGC ATTTTGCCAA	8880
GAAAATGATT TATTGGTTTG GTTAGGTGGT ATGTTTGAGT CTGGTGTTGG TCGCGCATTG	8940

AATTTGCAAT TTGCGTCACA GCCTACCTTT TCCTTTCCAG GTGATATTTT AGCAACGGAA	9000
CGCTATTTTT ATGAAGATAT CATCACGGAA CCTTTTATTT TAGAGCAAGG GACAATGACC	9060
GTGCTCAGG GACTTGGTAT TGGTGTTACA CTTTCCCAGA CGAATTTGCT AAAATACAGT	9120
CAGTATCAAA AAATCATGTA ATGATAACAA AATTGCGCCT ATGACTATGC AGAGAAGCAT	9180
AATCTAGGCG CAATTTTTTA TTTTCTAAT TGA CTGCTG CGGCTTCATC CAAGAGAACC	9240
GTTAAATTTG GATGGAGTTG AAGAATAGAG GCAGGATACG TTTCTGTGAC AGGTCCTTGG	9300
AGCACTTGTT TGACCATCTG AGCTTTTTGT TCACCATTTA CAATTAGGAC AAGATGCTTA	9360
ACTTTTCATGA TGCTGGCTGG CCCCATCGTA ACAAAGGTCA TGCCTGGCTG CATCATTTCA	9420
GGAACAAACC AAGGCTCTGA ACCAGTGACG GTCAC TTGGT AGGTTTCATT TTTAAAATGA	9480
GTGGTAGTGG GCATGTTCCC GCAAAAATGA CCATCTGCTC CTAAGCCAAT TAACATCAAA	9540
TCTAAGCCGC CTGCGTCAGC AATCCGTTGT TCTGTGTGTT GAAAGTTTTC AACAGTTAAT	9600
GGATGGATGT TTTGCTCGGA TATTTCCGCA GGTGTCAAAT ACAGCGTCCG CAAATCTGTC	9660
AGAGTAATGC CTTCCTTTTG ATGAGGAACG GGGATTTTCAT CAAAGTTATA ATAATGAACA	9720
TTATCAAAGT CAGATGAATT TTTCACAATC GAGGCTAATT TTTGGTAAAC AAGTGCTGGT	9780
GTTTTGCCAG CGGTAATTGA TAAATTCAC TCGTTATCTT GGCTCATCGT GCTTAGAAGT	9840
AACATTTTTG CCCATTCGCT CATCGTGTG AAATCTTGTT TAATTAAAAT TTTCATACTT	9900
GCTCACTCCA TTCTTTGAAT AAGCTAAGTA TAATTTGTGG GAACCATCCC ATGTCAAGCT	9960
CTTCACAAAA AGGATTTAAT TAACGGTAGG CAAAAGAAAG AGGAGCAGCC GTCAAAACGG	10020
CGACTGCCCC TCTTTTTTAA ACTTAAGATA ACCTACAACG AGTTAACTTT TTAGCTTTCT	10080
GTTCCACTTG CTCCAACGCT GGTCTAAACG TTCTTTTGAC CAACACTGTT TAACAAAGAA	10140
TTTTATTTTA TTGTTCTTTC AACCATTTTT CTAGTTGTAA ACGTTGCACT TTCATAGTAG	10200
CTGTGTGAGG GATTTCATCA AAAGCCATAA TGATAGGTTC GTTTAAATGT GGTAAATCGG	10260
CTACTTGTT CCACCAAGCG TCCCAATCCA TTTCTTCCCC AGGGACAACC GCTAAAATTG	10320
GCTGAGGTTT ATTTTCTTTG CCACGGACAA TAATCACTTC TTCTAAGAAA TCTAATGCAT	10380
CCAATAAATG ATCTTCAATA GCTAAATTAC TGTCGATGTT TTCAATTAGA TCTACTGGC	10440
GATCTTTCAA GAATAAGTGG CCGCGTTCAT CTAACATTCC GTAATCGCCG CTGTCCCACC	10500
ATTCGCCATA GACATTTTCT GCAAAGCGGG CATCTTCTTT GTAATATGTT AAGGCCCGTC	10560
CTTTTGATAA GAATTGGATA TGGCCGTCCG TCATCGCAGG TAAAACTTGT CCTTTTTCAT	10620
CGGTAATTCG AGCTTGTTG AAATCCTCTA AACCAACGCC CATGTCACGA GCATCTGAAT	10680
CTTTTAGGGA TTCTAAAGTA TGCGCTCGTA AGATCATTGG CCCGCATTCA CTTTGTCCAT	10740
AACTTGTAG GAAAATTGGT TCATTGGCAG CAGATGCTTT TAAGAAAGCC ACCATTGTAG	10800
CGTTGTTAAT TGCATCAAAA GTAGAATGAT AATAATGGAC ACTAGCGAAA GCTTCAGGTT	10860
TTTCTTTGGC CAAACGAACC CATTGAACGA AGTTATTGGG GTGTGTTTCT ACAGCGATTG	10920

GTTGATAGGT GCTGAACATT TCAGCGACTT TTTCGCTCGA AGCATTGCT AAAGGCATCA	10980
TAGGGAAGCC CATGGCCATT AAAGAAGACA CGCCAATGTT AAAACGAGAA TGAAGTGGCG	11040
AAATATGAAA GCCAACTAAT TTCTTTTCAG CAATTTTGT AAAAATGGTT TTTTGCCATT	11100
TTGTACGCCA ACCCATAGAG TGATTGGAAT GACAGATAAG TTTAGGAATA CCTGTTGTTC	11160
CAGAGGTGTG GGTTCATGTAG GCAATTCGT CTTTAGTCAG TTCGTCTTGA GGAAGTCCG	11220
TGGCATCAGC TTCTAAAAGA GAAGCAACCG AAAGTTTTTT ATTGGCAGAG CTATTACGAA	11280
CAGCTTGAAC TTTTTCAGCA GTCACATCAT CAAATAAAAT ATAAGGATCT TCTAAACGAT	11340
CAACGAATAC TTCAATTGTT TCAAAAGGTA AATGGTATGA AACCATTACA GGAACGGCTG	11400
CTAGATATGA GGCTGCAACA GCTAGTAAAT ACGTGTGCGAA TTTAGGACTT TTATAAATAA	11460
TAATTTTGTC GCCTTTTTTT ACACCTAAAG TGGCTAATTG ATATGCCCGT TTCAAAATTT	11520
CTTGATGGCT TTCACCATAA GTTGTCTTA AGCCAAGTGC AGGAAAAGCA GGCAAAGATT	11580
CATCAAAAT AATTGGAACA GTTGGTGTTC TTTCTGCGGC TTCTTGATAG TTTGTATAGA	11640
GGTTTAATGG TTGATAATCA AGTAAGTGTG TCAAGGGAAA GTCTCCTTTG TAAnATAGTT	11700
TGTGTACGAA TGTACAAGTT CACTTTATCC TATCATAGAA TGGGTCTTTT TCTACGTCAn	11760
AAAT	11764

## (2) INFORMATION FOR SEQ ID NO: 49:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11307 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

TCCAATATAA AGACACAACA GACAAACGaT GAGGAGTGTG AAwTAAtATG TTaACAATTA	60
TTATTGGGTT TATCTTTTGG ACAATGACAC TAATGTTAGG TTATCTAATT GGTGAAAGAG	120
AAGGCCGTAA ACATGAGTAA TTTAACAAAA CGTAAAAAAG ATTTATTGA AATGAAAAGC	180
GTTGTATTTA AAGATATTC AAAGCAACAA AGCGAAAAAG CACAAAAAAG AAAACGACTC	240
TTACAATAA TGAATCAATA TCCCGATTGG GCAAGTCAAA AAAATAAACT TATTATGCAG	300
GAAATTCAAG AATTAGGACA AGCAATCGGT AATTGGTCGA TGGATCAATC AAGACCCATC	360
CAATCCATCA AGGCCGCATC GTTTACAAAA AGCGAGTATC TCTATTTAAT TTGGCTCGGT	420
TATTCAGATG AAGCGATTCG TCACGGCTTA GACATGTCGA AAGAGTGTTA TTTTATTAT	480
CGATTAAACAC TTTTAAATGA ATAAAAGTAA AGGAGATTAA CCAATGCGTA CGTCAACATT	540
TAATTATATC AAAGATATTT TAGCAGACTT TTATAAACA GAAGAGTATA TCCGTCAACG	600
GGAAGAAGAA TTACGGCACC CTTATCAAGA AGCAGATTTA AATGCTGGTA TTAGAGGACA	660
AGGACTTCAC TCTGTAGTGA CCGAACGAAT GGCGATTACG ATAGCTATGG aTCGTCGTCT	720

GTGGAACCTTa	GAGAGAAATC	GAGACATTAT	CAAAAATTGT	TTAGCCGAAG	CGGaTGAACA	780
AACGCGCGTG	ATTATTGAAG	AACTATATAT	GAAAAACGG	CCCTCTTTAA	CATTAAATTGG	840
ACTTGCCCAG	CAATTATTTA	TTAGTAAAG	CCAAGCCTAT	AAATTAAGAA	ATCATTTCCTT	900
TGAAGCGGTG	GCGGATGAAC	TAGGGATGTA	AACATGGAAA	AAGCGTGGAA	TTTTTTCAGG	960
TGTCAACATG	GTAAATTAAT	AGTGTCGAAA	GAGATAGATA	AACGTGAGGC	AACCAAAAAA	1020
ATGAAGACAC	GGAATTCTAT	GATTTTGAAT	GCTTTCTTGT	GTCAGCTATG	AAGGAGCAGA	1080
AAATGCCGGC	TACTTTCAAG	ATCCTTCATT	TTGACTAGAA	GAGAGCCAAT	TTGTTAACCA	1140
ATCCTGAATT	TTTGAATGG	AAAGGTGGCG	CTAAAAATGA	ATGAAGCGGA	ACAAGAGTTA	1200
TATGAAGCCC	TTGTTGCAAT	CTGCCAGACG	TCaGGATTTT	TGTTGCTAGA	GGAAGTCCG	1260
ACAGATTTAC	CAGATCAGCC	ATTTGTTTAC	TTAGGTGATA	GTAAAGAATT	ACCTAAGCCA	1320
ACTAAATCAG	CTATTTTGGG	AGAGATTGAA	TTAATAATGC	ATGTTTATGG	TGCGTTATCT	1380
GAACGACAAC	AAATTTCTAC	AATTAAAGGA	ACGATTTTAC	GGCAGGCAAC	CAGTAACCTA	1440
AAACGAACGG	CTCATTTTAA	TTGGGGTATC	AAACATCAAG	AAGTCAAAGC	ACAAATGGTA	1500
AAAGATACCA	AACAAATGAA	AAAAACAATT	TGGCATGCTG	TACTACCATT	ACACATGCAA	1560
TTTTACTAGG	AGGAATTATC	AATGGGAGAA	GTTATGCAAG	GAAAAGACCG	TATTTTATTA	1620
GTTCGTCGCT	TGGATGAAGC	AGCGACAAAG	AAAGCAATGA	AACCCTTATT	TCAAATTGAA	1680
CATGAATGGG	AATTCTCACG	TGAATCGAGC	GGTACGCAA	CAAAAGATGG	CGTCGCGAAT	1740
GCTGTTTCTG	GTTTAGAAGT	TACGTTATCG	TTAAGCGGTT	TAGCCTCTCG	AGATGATGAA	1800
AATTTATACA	TGaAAGACGC	AgcGAaGATG	GCATCTTAAT	GGAATTTTGG	GATGTTGATT	1860
TAAAAGGTGA	AAAAAATGCG	GAAGGTAAAT	ATCCAGCAAT	TTATGCCCAA	GGTTATGTAA	1920
ATTCATGGAG	TTTACCAGCC	AATGTAGAAG	AATTAGTAGA	AATCGAAACA	GAAGCCTCTA	1980
TTAATGGCAA	GCCACAAGAT	GGCTTTGCAA	CAGTAGAAGC	AGATATTATT	GCAGAAGCAC	2040
AATATGCGTT	CCAAGATACC	GTTCCAGATA	AAGCACCACA	ACCTGGCGAA	TAATCAAAAA	2100
GTGTTGAATT	TTAGGAGGAT	AAAAAATGAA	TTTAGAGATT	AACGGAAAAA	CAATTGAAGT	2160
GAAATTTACG	ATTGGCGCGA	TTCGCGAATT	AGATAAACGT	TACCAAATTG	AAAATGGCGC	2220
TGCCAAATTC	GGCATGGGCA	TCAGTTCAGC	AATGATTTAT	TTACGCCAAT	ACAATCCAGT	2280
AATCTTAGTT	GACATCATGG	AAGCTTTACA	AAGTGGGCAA	TTAAAAATAG	GTAAGTCGGA	2340
AATTGAAGCA	TGGTTAATGA	CCCAAGATGT	CAAAAACTT	TCAGATGATT	TGCTTAAAGA	2400
AATGGGAAAG	CAACCTCTTA	CAAAACCAAT	GATCGATCAG	TTCAGCAAAG	AAGCGAAGAA	2460
AGCAGAAGCG	CAAssgACCA	ACTAATTAAA	ACGAGCGATG	ACGTGTATCA	CGACATCGCT	2520
CTTTCTGCTT	TTCGCTACTT	AGGCTGTCGT	TCATTTGAAG	AAGTGGATCA	GATGACCATG	2580
TCTGAATTTG	AATTACGAAT	GATTGCTTTT	AATTTAGCAG	AAGTAGATGA	AGAGCGGAAA	2640
AGGCACGAGC	TTGCCTACTT	AAATGTTAAA	GCGCaAGCGA	CAAACAAAAA	AGGAAAACCC	2700

GTTTTTGA	AAAG	TTTTATGAT	TATGAAAAAC	GAGTTGCTGA	AGTTCTGGCA	2760
GCTAACCAGC	CACAACGAAC	AAAATTAAAT	GAGCGGAAAA	AAACGCAACT	TGCCACTGTG	2820
GCAGAGCGTC	TACGCCGCTA	TCGAGAAGGG	AGGAGAGTAG	ATGGAGAATG	ACAAAGAAAA	2880
AACGCCGTTA	TCGGAGGCAA	AGAAAAGCCT	TGCAGGCGTC	CAACAAGCAT	TAAAAAGTAT	2940
GAGCGGTGAG	TATGCCTTAT	TAAGTGGATA	TTTAGGGAAA	ATTAGTGCGG	GTGTCAATCA	3000
GTCAGCCACG	GTCATGAACA	CATTTAAAAAC	CGTCATGCAA	CAATCTGGAG	AAACAGTGAA	3060
AAAAACAGGA	GACGAAACAG	CAAAGGCAGC	AGATCAAATG	AACACAGCGT	TAACAGATTC	3120
TGCTGAACAA	GCCGGTGaAG	CAGCTAAAAA	AGCGGGGAAA	GAAACCTCTG	ATGGCTTTAC	3180
TAATGCACAA	AATAATATGC	TGAGCTTTGG	GACGGCCATG	ACTAGTGCCG	TTTCCTTACC	3240
TATGCTGAAC	GTTTTAAAAA	CAGCTATGGG	CGTCGGTGCT	GGGGTCAGTG	GCGAATTTC	3300
AGGAATGCAA	GGACTGATTA	TGGCCAGTGC	AGGAGGGATT	TCTGATTCAT	TGCAAGGCGA	3360
GTTGCAAGGG	GCATTGACTC	AGATGAATCA	ATCATTTGAA	GCGGCGGCAC	AAGTGATTCA	3420
AAGCGTGATG	GCTCCAGGAA	TGGAAATTTT	GGTCAAGTG	GTTATTACAG	TCGTCAAAGG	3480
CATTACAGCT	TTGGTTAATT	TATTTATCAA	ATTACCAAAA	CCCGTCCAAG	TTTTTATTGT	3540
TGCCATTATG	GGCATTTTAG	CCGCCATTGG	GCCCATGTTG	ATTATGGTAA	CGATGGCTCA	3600
GCTAAAATtT	CAACAGTTTA	GTGCTGGTTT	GGCTCTTGTA	CAAGGAAACA	TTGGGAAGTT	3660
AGGTGGTGGC	TTATCAAAAC	TAAGTGCTAG	TTTTAGTGCC	TTAGGTGGAG	GACCATTAAT	3720
TTAATTGTA	GCAGCCGTTT	TAGCAGCGGT	AGCAGCGTTT	ATTTATTCT	ATAAAACCAA	3780
TGaAACaTTT	AGAAATAGTA	TCAATAGCTT	AGCTAGTGCC	ATTCmAGGaG	CTGTTTCAGC	3840
GGCGTTTGGC	AAATTGGTAG	GATTGCTACA	ACAGATCCAG	CCGGCCTTTC	AGCAAGTAAT	3900
GGCAGTTTTT	AAACAATTTT	TTGCAGTAGG	CTTAGAGAAA	ATGGCGACTA	TTTTTTCAAC	3960
AATTGGTCGT	GTGCTAGCAG	GCGTTTTTGC	CAGCGGTTTG	CAATTAGGTA	GTAACCTATT	4020
AGGGCAATTT	GGTGGCACCT	TTGACAAAGC	TGGTTTAGCG	GTTGGTCTTT	TGGTAAAAGT	4080
TCTGACAAAG	GTTGCACTGG	CTGCATTAGG	AATTCTGGGC	CGTTTGGTCT	AATTATTTCC	4140
TTGATTGTTT	CATTCGTGAC	GGCCTGGATG	AAAACCGGTG	ATTTGAGTGC	GGGTGGTATT	4200
ACCCAAAGTCT	TTGATAATTT	AGGTAACACG	ATTACATCGG	TTACAACAAT	GCTGGCAACT	4260
AATCTACCGA	AAGTTATACA	ACTTTTAC	ACAGTCTTAA	CCAGTATTCT	CGGGAAAATA	4320
ACAGAAGCTA	TTCCAAGCAT	CGTAACCGCG	TTATCTAGTT	TAATTACGTT	AATTGTTGGT	4380
GCGATCGTTG	CCAATTTGCC	AGTCTTAATT	GAAGCGGCAA	CACAAATTAT	TACTACGTTG	4440
ATTCAGGGGA	TTACAACAGT	CTTACCAATG	TTGATAGAAG	TTGGTTTGAG	CTTATTAATG	4500
ACTTTAGTTA	ATGCGATTGT	CACCGCCTTG	CCAACAATTA	CAACTGCAGC	GATTAATATC	4560
ATCACTACAT	TAGTGACAGC	TTTTGTCACA	GCGTTACCAA	TGCTAGTTAC	AGCAGGTGTT	4620
TCAATTATCA	CGGCCTTAGT	CAATGCATTT	GTTACTATGT	TACCGTTGAT	TTTGACTGCT	4680

GGTTTACAAA	TTTTGATGGC	ATTAATCACT	GGGATTATGA	CGATTTTACC	TCAGTTAATT	4740
CAATCAGCGC	TGACGATTAT	TCTAGCGTTA	GTGACAGCGT	TGATAGGTGC	CTTACCACAG	4800
ATTATCAGCG	CAGGTGTCAA	ATTGTTAATG	GCGTTAATTC	AAGGAATTAT	TTCGATTTTA	4860
CCAACCTTAG	TTGCGGCAGC	TATTACCTTA	ATTTTGACAT	TGGTAAATGC	CTTAATTGGT	4920
GCCTTGCCAC	AAaTCATCAG	CGCaGGCGTc	AAATTGCTAA	TGGCTTTGAT	CCAAGGGaTT	4980
ATTTCAATTT	TACCGCAACT	GGTTACTGCA	GCAATTACGC	TAATTACCGC	TTAATGGGT	5040
GCGTTAATCA	ATGCGTTGCC	ACAGTTGTTA	AGTGCTGGGA	TTCAACTGAT	TCAAGCCTTA	5100
ATTAATGGTG	TACTCAGTCT	ATTGGGTGCC	TTGCTGTCCG	CAGCAGGAAC	ATTAATCTCA	5160
CAAATGATCA	CGAAGATTGG	TTCTTATTTT	GGTCAACTGT	TAGCTTCGGG	CGGACAGTTA	5220
GTTGAAAATA	TCAAAAATGG	GGTTACCAAT	GCAGCCAATC	AGGTAAAAAA	TGCCATTGGT	5280
TCTGTAATTG	AAGGTGCTTG	GCAAGCAATC	CAAGGTTGGT	TTTCAAAATT	CACCGATGCC	5340
GGTGCGAATA	TTGTCGGCAT	GATTGCTGAT	GGAATTACAG	GCGCAATTGG	aAAAGCCAAA	5400
GAAGCAATCG	ATGGAGTTGT	CAGTAAAATT	CGTAACTTTT	TACCATTTTC	ACCAGCAAAA	5460
GAAGGTCCCT	TATCTGATTT	GCATAAATTG	AATTTTCGGCG	GCACGATTGC	CACGGGGATT	5520
TATGCAGGCG	AAACAGCCGT	TAGTAGAGCA	ATGGCTTCTA	TTTAGATTTT	ACCGCTGTTA	5580
AATGATTTTG	CCTTGGACTT	AGCTGGTCGA	GGAACTTCA	CGGCAACGAT	TGACCATCGT	5640
TTAGAAAATG	ATGCATACAA	TCGACCATTA	TTGTGACAG	TAGAGTCAAC	GTTAGATGGA	5700
AAAGTTGTCG	CAGCAACTAC	GGCGCCTTAT	TTAGCAACAG	AGTTACAACG	ACAACAAGTG	5760
AAACAAAATA	ACCGCTTAGG	AAGGAGAGGA	TAACATGTAT	AAATTTGTTG	ATACCAATCA	5820
AGCAACTCAT	TCAACGCCTC	TTCTTCAGA	AGCGTTGAAT	TTTAACGGCC	AATTTTTAGA	5880
AAAAGTCATC	CCTGGCTATC	AAACATTATC	AGTTTCAGGA	CGAGAATTAG	TTCCAAGCGA	5940
AATTGAAAGC	TATCAATTAG	GGATTCTGTA	TGGTAAACGT	CACGTTTATG	CGCGAATTCC	6000
AGAACGAGAA	TTAACAGTCA	AATATCGCCT	TTCAGCTGTG	AATAATGAAG	CATTTGAGAA	6060
TGCATTTAAT	CATTTAAACG	TTGCTTTGTT	TACGGAAAAA	GACGTTtCTA	TTTGGTTTAA	6120
CGATGAACCG	GAAATGCTGT	GGTTTGCCAG	TAAGTCTTCA	GTGAGTGATG	TACCCGAAGt	6180
GTTAACCAAG	TAACAGGCAC	CTTACTTTTA	TTGCTTTCTG	ATCCGTATAA	ATACACACGG	6240
AGkATgGCGA	CTAGTGTGAT	GTGGGGTTCTG	CCAACCATTA	CATTTCAAGC	GAATTACTTA	6300
ATGGGGAATA	CAGGCTCAGG	TGCATTTGAT	TTTCCAATTT	TAATTGAAGG	CGGGGCTTAT	6360
TGGGGATCAA	CCATGATTAC	CTTCAAAAAT	CGGGCTTACA	CGATGGGGGA	TTTAGGCAAA	6420
GAAGTTCGGC	CAATTGAAAT	TTATCCTACG	GTTGAAGGAT	TAAAAGTCAA	ACCGACCATT	6480
ATTTTAACAG	GAACCGGACG	TGGTGTGTTG	ATTAAAACAC	GGAACGATAC	AATTAACTTA	6540
GGAGACTTTG	ATCGTTCGGA	AATTATTATC	GATACTGAAA	ATTTTTATCT	GACaAAAAAT	6600
GGTGCACCGA	TGATTCGACC	AATGAACGAT	TTTTATCTAT	ATCCCAATGA	ACCGCTGTAT	6660

ATTCAAGCCA AAGATAGCGA CTTCCGCTTG ACGATTTCGCT ATCCTAACCG ATTTGTGTAG	6720
GAGGGTGATT AAATGTTAAT GCGCTGGAT TTGAAAAGAA CATATACGGC AATCTTGAT	6780
AATGCCATATC AAGTCAGTTA TGAAAAAATA GAGAACAAAA TTGGCAGTTT AGATTTTACC	6840
ATGCCACTAG ATGATCTTAA AAATGAATTT ATTGCAGAAA TGCAATGGGT GGAAGTACC	6900
GACAATGAGA ATGAATAATAT TGGTTTATAT CGCGTGATGC CAACCACAAT TAAGAAAGAT	6960
GCGAACATA ATCAAATTCA TTAATCTGCC ACAGAAGCAT TATGTACCTT AGGTGATACT	7020
GTCCTTTTGT GTTGTACGA AATTAAAAAC AAAACAACGA AAGAGGCCAT TCAATTTTGA	7080
TTGAATAAAC AAAAAACAAA GCATTGGGTC CTAATAAAAT GTGATTTTTC AAGGAAATTA	7140
ACCTATAAAT GGGAAATGA AAATGGGTTA GTTGAGCCTT TATTTAGCAT CCCAGCCGAT	7200
TTCGAAGaGG AATATCTTTG GCAATGGAAT ACAGAGGTCT ATCCTTTTGA ACTTTCATTA	7260
GTCAAACCGC CAACAGAACC AGTTGCGCGA ATTCAAGAAG GTTACAACAT GCAAGGATTT	7320
GAAATAGAAC GTAATCCCAA GATGCTAATC AATCGGATTT ATCCATTAGG TTCAGGCGAA	7380
GGTGTTAACA AAGTCAATAT TCGCTCGGTC AATCAAGGGG TTCCGTATTT AGAGAACAAG	7440
CCCGAATTG ACCGCTATGG TTTATTGGAG TCAATTTGGG TGGAACAGCG TTTTCTGAT	7500
CCCAAGGCAT TAAAGGAAAA TGCTTTGCGA ATGTTAGAAG AATGGACCAA GCCACAAGTT	7560
TCTTGGGTAG TGAATGACG TGATTTAATT AAATTAACAG ATCAACCTTT GGCAATCGAT	7620
CGTTTGCGGT TGGGCACGGT TATCATGATT AATACGAATG AATTGAGGAG TGTCAACCTT	7680
CGTATTAAAA AAGAAAGCAA AAAAGATGTC TTTGGTGCCC CCCAAGACAT TCAGCTAGAG	7740
TTGGGAAACC TGCAAGAAAC AATTCATAGT ACCATGACAG CTTTCAGTCG GAAACAAGAG	7800
ATTAACGAAA CTTACGCACA AGGGGCGACG AACTTTTAA ATCGTTCAAT ACAAGGAGAA	7860
CTTAGCAAGA CACAGCCAGT GGAGCTGAAT TTATACTTTG ACGAGGACAT TCTTTATATA	7920
AACACCGCAG AATTAACGTT CAAGGCAACT GCTAAAGGAC CTTTCGATTC TGTAACGAAC	7980
ATTGATTTGG TAGTGGATGG CAAAAAATTA CCCCAACTAT CATTGCAACA ACAACGGCTA	8040
AACATTTTGA GTTATTTACG AAAAAACAACA GATGAAAAAA TCGAACGCGG CAATCACACG	8100
CTTCAATTTT TCTCTCATCA GCCACTATGG TTGGATGCTT CGGTCATCTG TCGTGTGTAT	8160
ATTCAATCCC AATTGGGTGG CCAGTTTAA TAAATAATG AAAACTAGAG GAGTGTGACG	8220
AAATGTCAGT AGAACATATT GAAGAATTAG ATACCCTGAA TCAAGGTCGC CTAAATCA	8280
ATGCAATCTT GGATCAGTCG AATGCATCAG CTGAGAAAGT AGATGCTTAC CAAGTCCAGT	8340
TAACGAATGG AATTTCTGAA GCGAAAAACA TAGCAGATGA AGCTGGCAAA GAAGCCGTAC	8400
AAATTGCCAC CGATGCAGGC AATCAAGCaA aTGAAACAGC CAACCAAGCG aTGaACaATG	8460
CCAAAACAGC AaTCATGATT GCAGGaAATG CAGTTTCAAC GGCAAATAAT AATAACAAG	8520
AATTTGATAC TTTACGAAAT GATTTCGATC AATTAGTAGC AGAAGCGGGT GATAGTAATC	8580
CAGAAATTGT CCAAGCACGT ACAGATACAC AAGGCATCAA ACAAGCTACC TTAGCGAATC	8640



GTCTTCAAAT TGATTTGAAT GACCGTATGA CAAAAGCAGA CGGTATTTCT TTATTGGCTA	8700
AGCCAACTAC TGTCAAAATG AAGTTAGACT TTAACGGTAA AACGGCCGGC AATACAGCCA	8760
CCAATGCAAA CAGTTATTCC ACTGATTTTA CGGCAAAAAT TCTTAAGAAG CCAACAGAAG	8820
TTTGGGAGGA AGTTTCCCAA GCGGACTACA ATAAAATGGC CAGCCGTGAT GATGAGGGCG	8880
TGAAAACAGG TTCCACCCAA AGCGGTGTGA TTCCGCAACA ATTAGCGGCC TTCAATCTCG	8940
TTGAAGCCGC TAAAAAATTA ATTCCACAAA TGTTTGAAAC AGTCACAAC TACGAGGCGG	9000
TGGCATTAT TCGCCAGAAC GTTCAATTTT TTACGATTAA TCAACGTGTG AAAGCCGCTG	9060
CGCCCAATAA TCAAACGATT AAAATCGCTA CGTATTTACC AACTACGGAT AATTGGGTAA	9120
CTCAAATCCA AGAATCAGCA AAAGAGTTTG GCGATTTTTC AATTCAAATC AATGATCAGA	9180
ATTTTATCAC AGATGAAGGT TTCATTTATT TAATGAGCTA TACAGATTCA TCGAATGGGG	9240
TAACGCCAGC TAGCTTAGAA GTTGATTACG TGGGGCTTCA TATTGGTCTG TCTGTTGATG	9300
CCCAAGCGGT TTTAGCGAAG AGTGGTTTTG TTCAAGCAGA GCAACTCAAT ACCCATATGG	9360
AAAATCAAGA TAACCCGCAC CAAGTAACCG CTGAACAAGT GGGGCTAGGC AATGTAGAAA	9420
ATTATGGnTT CGCATCAGAC AGCGAAGCAG TCGCGGGAAC TTTAACGAGT AAATATATGC	9480
ACCCGAAAAA CGTTGCGGAA GCGATTAAAG GTCAAGCTGT GACACAAACA GGTGATCAAG	9540
AGATTGCTGG GGTGAAGAAT TTTGTAATA TGCCAACCGT CAATGGTGTG CCTTtGGAAT	9600
CCTCTAGAAT GGCCATTTAT GAAGCTAnTG GAGTTGGTGA AGyCGAGGCA AAATATCAGg	9660
CGGCCTTTAA TAAGGAtAAT ATGAAATTTG TATTAATTAG GGTAGGAAAT CGTGTGATG	9720
CATTTGTAAG ATGTAATTTG AGTGATCCaA CGAAATTGAA TAATAATTTG GTTAAAGTGT	9780
TTACTGTTCC AACAGGATAT ACATTaTCGA CGAAGATTAC AAAGGGAATA TGAATyTGG	9840
CGTTAACTGC TATGCAATAT ACATTCCCTC AACCGAATTG TGCTGGTTTA TATGAGATGG	9900
GAAATCAAGG AATTCTTTTT GGTGCTAACC GTGCTGGAAA TATTTACCTA CAAGGAAGTT	9960
GGTACACGGA CGATCCGTTT CCAACAAAAT AACAAAGCTAG TTTAGGAGTC TTTTATGGA	10020
ACGTTATCTC AACACAATAA CAATGCTTTT AAGCATTTTC GGTGGGATTG TCGTACGTTT	10080
AFTAGGCGGA TTAGATCAAT TGTTGGATGT CTTCTCTTT TTAATTATTG TCGATTTCAT	10140
CACAGGTTGG ATTAAGGCAA TCGCCACAAA AGAATTGTCC AGTCGGATTG GTATGCTCGG	10200
AATTGCGAAA AAAGTGACGA TGTTATTTGT GGTGCGGTA GCGGTTCTGT TTGAAAAAGT	10260
TGTGGGGAAC AATTTGCCAA TTCGGGAAAT GGTCTGATT TTTTACATTG CGAACGAAGG	10320
ACTTTCTTTT TTTGAAAACA TTGCGACCTT TATTCTATG CCGAAAAAGT TAAAAGAGTT	10380
ATTTATTAG TTAATAATAA AAGATGATTA AGTAGAAGTG GTCGGGACAA ACGTAGAACT	10440
TTCGcTGATT GCCGAAGAAA TTAATTCTGT CCCGCCATTT ATCTGCAGGT TTAAGCCGTG	10500
GAAGGGAAGT TATTTTGAAT TTCCTTTTCAT GGCTTTTTTA AGAAAGGAGC ATGCTATGTT	10560
TAAAAAATTA ATGATTCAAC TTGCTTTAGT GATTGGTTTA AGTTTAACGA TTCCGATGAC	10620

GGcTrCGCT TACACCATCG AAGCGGATCC AATCAACTTT ACTTATTTTC CCGGCTCTGC 10680  
 AAGCAATGAA TTAATTGTTT TACATGAATC TGGAAACGAG CGGAACCTAG GACCACACAG 10740  
 TTTAGACAAT GAAGTGGCCT ATATGAAACG AAATTGGTCA AATGCTTATG TCTCATATTT 10800  
 TGTCCGATCT GGTGGACGAG TGAAaCAATT AGCTCCTGCT GGCCAAATTC AATATGGCGC 10860  
 AGgTTCTTTA GCTAATCAAA AAGCCTATGC GCAAATCGAA TTGGCTCGAA CGAATAATGC 10920  
 GGCGACATTT AAAAAAGATT ATGCTGCCTA TGTTAATTG GCCCGTGATT TGGCTCAGAA 10980  
 CATTGGTGCT GaTTTTTCTC TGGaCGATGG AACAGGTTAT GGCaTAGTCa CTCATGATTG 11040  
 GaTTACAAAA AATTGGTGGG GAGATCATAC AGATCCTTAT GGtTATTTAG CGCGTgGGGG 11100  
 AttAGTAAAG CGCatTGGcA CmAGATTTaC aAcGGGCGTT TCyGrAACAg GTGaGACTGG 11160  
 TCATTATTCA GCCAGGTAAA CCTAATGCGC CAAATATCA AGTAGGACAA GCAATTCGTT 11220  
 tCACTTCAAT CTATCCAACA CCAGATGCTT TAATCAATGA ACATCTATnA GCAGAGGCAC 11280  
 TTTGGACACA AGTAGGAACA ATTACAG 11307

## (2) INFORMATION FOR SEQ ID NO: 50:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2866 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

AAAGCACAAA TGCAACAAGT TGTGATAGCT ATTCAAATA CGGAAGCCTT TCAGTTATTG 60  
 AATGTAATGA AGGTTATAGG AGATGAGTAA ACATGTATGA AGGTTTGTTA AAACAGTATC 120  
 AAGCGTATTT GCCCGTAACA GAAAAACAC CAATGATTTC TTTAGCCGAA GGAAATACGC 180  
 CACTGATTCC TTTACCTAAT TTATCCAAG AATTAGGTAT TCAATTATAT GGAAAATACG 240  
 AAGGCTTAAA TCCCACTGGC TCATTCAAAG ATCGCGGGAT GGTGATGGCT GTAGCAAAAG 300  
 CGGTAGAAGA AGGTGCCAAA GCGATTGTTT GTGCTTCTAC AGGAAATACT AGTGCCGCCG 360  
 CTGCTGCGTA TGCCACAAGA GCAGGCATTA AAGCGTATGT AGTGATTCCA GAAGGGAAAA 420  
 TCGCACTGGG AAAATTGGCG CAAgcTATTA TGTATGGGGC GGACATTATT TCGATTCCAG 480  
 GTAACCTTGA TGAAGCATTA AAAGCGGTGC GAGAGATTGC GAAAAGTGAG GCAGTTGCGC 540  
 TAGTTAATTC AGTGAATCCT TATCGCTTAG AAGGGCAAAA AACAGCAGCT TTTGAAGTTT 600  
 GTGAACAATT AGGTCAAGCG CCCGATGTCT TAGCGATTCC AGTTGGAAAT GCCGGCAATA 660  
 TTTCCGCTTA CTGGAAAGGC TTAAAGAAT GGCATGAAAA ACAAGGAACA ACTTTACCAA 720  
 GAATGCATGG TTTTGAAGCA GAAGGTGCCG CAGCGATCGT GAAAGGCCAA GTGATTGAGC 780  
 AACCTGAAAC CGTGGCAACA GCGATCCGGA TTGGGAATCC TGCTAGTTGG CAATTAGCAG 840  
 AACAGGCACG GGATGAATCC GGTGGTTTCA TTGATGCCGT GACAGATCAA GAAATTTTAA 900  
 CGGCTTACCG AAAAATTGCT GCTCAAGATG GGGTTTTCAT TGAACCTGGT TCGGCTGCTT 960

CATTAGCTGG AGTTATTCAA CATGTTAAAA GCGGTAAAAT CAAAGCAGGT GAAACCGTCG	1020
TGGCAGTCTT TACTGGTAAC GGAAGTAAAG ATCCTGATAC AGCAATGGAG ACAGAAGTAG	1080
CTATTTCAAA AATGAGTGAC GTGGAAGaAA TGCGTCTGcA CTTACgAAAA GGAGTCGCAA	1140
CACTATGaAA ATAAGAGTAC CTGCCaCtAG TGCCaACTTA GGACCTGGCT TTGATTCATG	1200
TGGCATTGCA TTATCCGCTT ATTTAACAAT CAACGTTTTA GGTGAAAGTG AATTTTGGGA	1260
AATTCAACAT ACTCTGGGAG AAGAGATTTC TACTAATGAA GAGAATTTAT TAATTCAAAC	1320
GGCATTGAAA ATTGCACCAG AATTAACACC AAAAGTTATT CGGATGGTTT CTGACATTCC	1380
CCTTGCAAGA GGAAGTAGCA GTAGCTCAAG TGTGATTGTA GCAGGCATTG AATTAGCTAA	1440
CCGACTAGCA CATTTAAATT TGTCCCCTAA AGAAAAAGTC CGTTTGGCCA CAGAGATGGA	1500
AGGACATCCG GATAATGTGG CCCAGCTAT TTTAGGAGAT TTCGTGGTGG CTAGCCATGT	1560
AGAAAACCAA GTCTACCATG TCAAACATCA CTTTcCAATG tGTGATGtCA TTGcATTTAT	1620
TCCAGAGGAA CCATTATTTA CCGAAAAAAG TCGCGCTGTT TTGCCCCGAAA AATTGGCTTA	1680
TAAAGAAGCG GTCGCGGCCA GCTCAATTGC CAATGTTATG ATTGCCGCAA TTTTAAACGG	1740
GGATTTACCT TTAGCAGGCA AAATGATGGA ACAAGATAAA TGGCATGAAA CGTATCGTCG	1800
TTCATTGGTT CCTCACTTAA AAGAAATTCG TCGTTTAACT CAGCAAAAAG GTGCTTATGG	1860
CTCTTTCTTA AGCGGTGCTG GTCCAACCTGT GCTAATCTTG TCGCCAGAAG AACGAACCAA	1920
CGAAATTGTT CAAAGCTtGG AGAAATTATC TACCAAAGCG AGCATCCAAA TTTTCAATAT	1980
TGATCAAGAA GGTGTTCAAG TCTTTTAAAA AAAACGAGAA AAAATCACTA GTGATTTTTT	2040
CTCGTTTTTT TTAATCCTCG GATGTTTCCG TGGAATAATA ACTATGATCT TCTAAGTTCA	2100
AAGCATTAAG AATCATTGAG GCAGTTTCTT CAATGAAAG AGAAGCGACA TCAATAACGA	2160
TACAACCTAA TTTTGGATAA AGATCATTAG CAAAAGCCAG TTCTCGCCGA ATTTTTTCAA	2220
TATCTGAATA AGAAGTATTT TCAGGCAATC CATATGCGCG CATTCGTTCT TTACGAATCC	2280
CGTTTAAAAC ATTGGGATCA TTGGTTAAGC CAACAATTTT TTTAGGGTCC ATTTGGAATA	2340
ATTGTTTTCG TAAATGTGCT TCAGGAATTA GTGGCAAATT AGCAACTTTT AAGTTCTTGT	2400
TGGCTAAAAA GAGACTTAGC GGTGTTTTAG AAGTTCGGA AACGCCAAGA AGTAACACAT	2460
CCGCTTCTAA AAAGCCTCGT GGGTCCTTGC CATCGTCATA TTTGACAGCA AATTCCATTG	2520
CTTCAATTCG TTTAAAATAA TTTTCATTCA GATGATGTAA GGCACCTGGT TCCCGCATGG	2580
GCGCGACACC TGTTAATCGT TCGATTCCG CAACTGGTGG TGTCAAAATG TCAAAATGGA	2640
ACAATTGATG CTCATTGAAA AATTGATTAG CAATCGCCAC GAGATGATCA TTAATGATms	2700
tATGCAGTAC CATGGCCTCT GCTTTTTTGG CATCTTCTAA CGCTTGCACC AGTTTATCTT	2760
CGTCTTTAAC GAATGTTCGT CTAATCAAAG TAAATCAAC AGTGGGGTAC TGGGCCATTG	2820
AAGCAGCCGC CAATTTTGAA GCGGTCTCGC CAGCAGAATC TGAGAT	2866

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1663 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

GATTnAAAn nAATTAATTG AAAATACCCG CTGsCatGaC ACGGGAAGGG aTTACGAAAG	60
AAATTCCAAT GGACGAAATT GTTCCAGGAG ACATTGTAAC TTTAGCTACA GGAGaCATGA	120
TTCCTGCAGA TGCCGTGTG ATTTGGACCa AAGATTTATT CGTTAATCAA TCgTCGCTAA	180
CAGGCGAATC CATGcCyswC GAAAAATTG TGGATGCAGG TGTGGATCGT CAACAAACAG	240
AGGTTTCTGC ATTAGATATG CAAGACCTAG TGTtTATGGG CACAGATGTC TTGAGTGGAC	300
AAGGAAAAGC GATTATTTTA AAAACTGGGC AACATACATT TTTTGGTGAT ATTGCTAAAA	360
ATGCTACAAC ACAACGAGGA AAAACTAGTT TTGACTTAGG CTTAGCAAAA GTCAGTAAGT	420
TTTTATTACG AATGGTAATG ATTTTGTTC CAATCGTTTT CTTGATTAAT GGTTTAACTA	480
AAGGTGCTTG GGGCGAAGCG TTTTCTTCG CGATTGCCGT TGCAGTTGGT TTGACACCAG	540
AAATGTTGCC AATGATTGTT ACGAGTAACT TAGCCAAGGG TGCATTATCC TTATCGAAAC	600
ATAAGTCAT TGTTAAAGAA TTAGCGGCTA TCCAAATTT AGGTGGTATG GATGTGCTCT	660
GTACAGATAA GACTGGAACG ATTACCGAAG ATCGTGTAGT TTTGGTTCAA CATTGAATC	720
CATTAGGTGA TTTGAGTGAC GAGGTTCTAA ATTTAGCCTA CTTAAATTCT TCTTATCAAA	780
CAGGTTGGAA AAATTTAATG GATATTGCCG TTATTAATTT TtATGAAGAG CATCAATGGA	840
AAACGCCGTT TAAAAATGTC ACTAAAATTG ATGAAATTCC TTTGATTTT TCTAGACGCC	900
GCTTAACGGT TGTCGTGAAT GCAGATGATC ATCAATTGAT GATTACGAAA GCGCCCGTGG	960
AAGAAATGGA AGAAGTGTGC ACGCATGCGC AAATTAATGG CGAAATTGTA CCATTAAGCA	1020
CTGCTGTACG AGAAGAACTA CGCCGAGTGA ATGTTCAAAT GAACAAACAA GGGATGCGTG	1080
TGTTAGCGGT AGCTGTAAAA AAAGATGTCC ATAAAGAAGC AGTTTATTCA GTTGAAGATG	1140
AAAAAGAGAT GACATTGATT GGGTTTATGG GTTTTCTTGA TCCAGCAAAA GAGTCAGCCG	1200
TCAGCGCTAT CCGCTCCTTA CATGAACATG GTGTCAACGT GAAAGTGCTA ACGGGCGATA	1260
ATGACATTGT CGCCAAAAAA GTCTGTAAAG ATGTAGGCAT TGAAGTTTCC CATGTGTTAT	1320
TAGGTTCTCA AATTGAAGCA ATGACAsnTG AAGAGTTGCG GCGCAAGTA GAAGAACTA	1380
ATTTATTTGC CAAATTAAT CCGATGCAAA ATCTAAAATT ATTGAATTGC TACAAGCAAA	1440
AGGTCATACG GTTGGTTTTA TGGGGGATGG CATTAAATGAT GCACCGGCAT TACGAAAAGC	1500
TGATGTGGGC ATTTAGTAG ATACTGCAGC TGATATTACT AAAGATGCTA GCTCGATTAT	1560
TTTACTGGAA AAAAGTTTAA ATGTACTGGA ATCTGGTGTG ATTGAAGGAC GAAAAGTATT	1620
TAGCAATATG ATGAAATATA TTAAATTAC CATCAGTTCT AAT	1663

## (2) INFORMATION FOR SEQ ID NO: 52:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5663 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

```

CTTAATGGnG ATGGnAAGGG GATATTATTG GTTATTTTTT GAGnATATAT GATATCCTTT      60
ATTATGTAGG AATCnTATTA TATGGsAGGm ACGAGCGATG AAAGTTATTG TTTTAGGATC      120
ATCACATGGA GGTTATGAAG CGGTAGAGGA ATTACTAAAT CTTCATCCTG ATGCAGAAAT      180
TCAATGGTAT GAGAAAGGTG ATTTTATCTC ATTCTTGTCT TGTGGCATGC AGTTGTACCT      240
AGAAGGAAAA GTGAAAGATG TTAATTCTGT TCGCTATATG ACTGGCGAAA AAATGGAGAG      300
CCGTGGTGTA AATGTCTTTT CTAATACTGA AATTACAGCG ATTCAACCAA AAGAACATCA      360
AGTGACAGTG AAAGATTTAG TGTCAGGTGA AGAACGTGTT GAAAATTATG ATAAATTAAT      420
CATCAGTCCC GGAGCTGTCC CATTTGAATT AGATATTCCA GGTAAGATT TGGATAATAT      480
TTACTTGATG CGTGGTCGTC AATGGGCCAT TAAATTAAAA CAAAAACAG TAGATCCAGA      540
AGTCAATAAT GTGGTTGTGA TTGGTAGTGG TTATATTGGG ATTGAAGCTG CCGAAGCATT      600
TGCAAAAGCC GGCAAAAAGG TTAAGTTTAT TGACATTTTA GATCGTCCAT TAGGGGTATA      660
TCTAGATAAA GAATTTACAG ATGTTTTAAC AGAAGAGATG GAAGCTAATA ATATTACCAT      720
TGCAACTGGT GAAACAGTTG AACGTTACGA AGGCGACGGT CGTGTGCAA AAATCGTTAC      780
AGATAAAAT GCGTACGATG CTGATTTGGT CGTTGTAGCG GTTGGTGTCC GTCCAAACAC      840
TGCTTGTTA AAAGGTACCT TGGAATTACA TCCGAATGGC CTAATCAAGA CGGATGAATA      900
CATGCGGACA AGTGAGCCGG ATGTATTTGC AGTAGGGGAT GCTACGTAA TTAATACAA      960
TCCTGCAGAC ACAGAAGTAA ATATTGCCTT AGCAACGAAT GCTCGTAAAC AAGGTCGCTT     1020
TGCTGTGAAA AACCTAGAGG AACCAGTTAA ACCTTTCCCT GGTGTTCAAG GATCTTCTGG     1080
CTTGCCGTC TTTGATTATA AATTGCTTC AACAGGGATT AACGAAGTCA TGGCTCAAAA     1140
ATTAGGAAAA GAAACAAAAG CGGTGACAGT AGTAGAAGAC TACTTGATGG ACTTTAATCC     1200
AGACAAACAA AAAGCTTGGT TTAATTAGT GTATGATCCT GAAACAACAC AAATTTTAGG     1260
CGCTCAATTA ATGTCGAAAG CAGATTTAAC TGCAAACATT AATGCTATTT CATTAGCGAT     1320
TCAAGCCAAA ATGACGATTG AAGACTTAGC CTATGCGGAC TTCTTCTTCC AACCAGCGTT     1380
TGACAAACCT TGGAATATTA TTAATACAGC GGCTTTAGAA GCGGTGAAAC aAGAACGTTA     1440
ATAGAAGAGA mAATAAAACC CATGAGTATT TATAGTACTC GTGGGTTTTa TTTTATTTT     1500
GTATTCAATT GATTTTTCAG CTCTTGTGG ACACGAGTTA ATTCTTGTTT ATCCACACGT     1560
TGATAGGAAA CACCATCTTG CATAAAACCA GTACCTTGAA GTTGGTCTTG TTCACTTTA     1620
CCAAAGGCAC TCGATAATC TAAGGCAATT TTTTTCATGT CATCAAACT TAAATCTGTC     1680

```

TTCATATTAT CAGAAACAGC TGTTAAAATT TCTTGaTAGT TGCTTACGCT GTTAAGACTT	1740
AAGACTTTTT GGACGATGCC TTCAATCACT TTTCTTTGAC GTTCTTGGCG GCCGTAGTCA	1800
CCATTAGGGT CTTCGTAACG CATTCTTGAA TAGGAGAGTG CTTGTTCAAC ATCCAATGAA	1860
ATTTTACCAA TCGTAAAATC ATATCCGTCT TGAGAAAAAG TCAGATTATT GTTCACTTCG	1920
ATTCCGCCAA CCGCGTTGAC TAATTCTTTT AAACCAGCCA TATTAATTGA AACATAATGA	1980
TTAATAGGTA TGTTTAAATA GTTTTCAACT GTGTCCATTG CTAAAGATGC GCCACCAAAA	2040
GCATAGGCGT GATTCAATTT ATCTTGTTTT CCTTGACCTG GAATATCAAC ATAGGTATCG	2100
CGAGCAAGAC TGACTAAGGT TGTTTGCTTG TCACGAGGAT TAACTGTTGC AACAAATTGTT	2160
GATCCGAAC GACCTTGCTC GACACGCCCA TCATCGCCTG TATCAATCCC TAATAATAAA	2220
ACAGAAAAAG GCTCCTTATT GTTTAAATTG ACCTGACTTT TTTTAGATCG TTCTACTGTT	2280
TCATAGGTTT TATCCATTGA TTTAGAAACA TCCCAATAAA GTTTTGCTCC CATtCCAaCA	2340
ACTGCAaGAA aTAGAmCCAa GATAATTCCA ArGATAATGG CAAAAATTTT CTTCCCTTTT	2400
GACATATTTT yCCTCCAATA ATTTATCACA GGTGTGTGATT TTTTCTTAT AGGTTAATCT	2460
TACTAATTTT TAGTGAAGAA GTATATAATA ATTGTTGAAA GACTATAAAA CTTTAAGGTA	2520
TTTTGCATTA TTTTGCTTA ATCTTTTGAA ATATCTCAAT TCTATTAAAA TATGCGAATT	2580
AGTATATTAC ACGTTTCCAA TGTGTTTCAT TGTTGATTAA TTAGAATTTT AGTGGTATTT	2640
TTGTATTTGT AATTAGTACA GATTTTAATT GGAGGTATAA CAGTGAGTAA AAAAGGATCA	2700
GATATCATAG TAGAAACtTt AATCAATCAC GAAGTGCCCT ATATTTTGG GATTCCCGGT	2760
GCAAAAATTG ATGGGGtTTT TGATGCATTA GTtGACAAGG TCCAGAATTA ATTTTAGCTC	2820
GACATGAmCA AAaCGTGCC TtTATGGCAC nAGGGaTTGG TCGTTTGACG GGAGAGCCTG	2880
GCTTAGTTAT TGCAACTAGT GGACCTGGTG CCAGCAATTT AGCAACTGGG TTAGTAACTG	2940
CGACTGCAGA AGGGGACCCT GTTTTAGCAT TGGCTGGACA GGTGAAACGG TCCGATTTGT	3000
CAAAATTAAC TCACCAAAGC ATGAATAATG CTGCATTATT TAGTCCAATA ACAAATTCa	3060
GTACAGAAAT TCAAGATCCA GAAACATTAT CTGAAAATAT TGCCAATGCT TACCGTATCG	3120
CCAAAACAGC GAAAAAAGGC GCAACATTCT TATCGATCCC ACAAGATGTC ACAGATAGTC	3180
CGGTGACAGG CGAAgcAATT AAGCCTTTGT CGGCTCCGAA ACTGGGCCAT GCATCAGCAG	3240
GAACAATCGC AGCTTTAGCT GAACGAATCA AGCAAGCGAA ATTACCTGTT TTACTTTTAG	3300
GAATGCGGGC CTCTGCACGA AAAGTTACCG CAGCCATTCTG TGAATTGGTT GCTATAACTG	3360
GTTTGCCAGT CGTGGAACG TTCCAAGGAG CAGGCATTAT TTCAAAAGAA TTAGAAGAAC	3420
ATTTCTTTGG GCGTGTTGGT TTATTTGAA ACCAACCTGG GGACCGTTTG TTA AAAAGAA	3480
GTGATTTAGT GATTGCGATT GGTATGACC CAATTGAATA TGAGGCACGT AATTGGAACG	3540
CAGAAAAAGA TGCGCGCGTT ATTGTTATTG ATGAAGTTCA AATGGAAATT GATCAATATA	3600
TGCAACCGGA AGAAGAGCTA ATTGGTGATA TGTCTAAAAA TATTTTGAAA TTATCAGAAG	3660

CATTTAGTGA ACCAATATTA ACAGAAGATG CGCAAGATTA TTTAGAAACG CTTCAAGAAA	3720
AATTAACGAT TAAAGAAGTA AAAACAAGTA CAATTGAAAA TCGATTACAT CCATTAGAAA	3780
TTGTTCAAAC GCTTCAAGAA AAAACCACGA ATGAAATGAC AGTTACTGTT GACGTAGGGA	3840
GTCACATACAT TTGGATGGCT CGCCACTTTA GAAGTTATGA ACCGCGTCAT TTATTATTTA	3900
GTAACGGCAT GCAAACATTA GGTGTCGCTT TGCCTTGGGC AATCTCAGCT GCACTAGTTC	3960
GTCCAAATAC ACAAATTATC TCTGTTTCTG GGGACGGCGG TTTCCTATTT TCGGCGCAAG	4020
AATTAGAAAC AGCGGTTCGT TTAAAGCAAA ATATCGTGCA TATTATTTGG AATGATGGTA	4080
GTTATGATAT GGTAAATTC CAAGAAGAAA TGAAATATAA TCGATCCTCT GGCGTAGATT	4140
TTGGTCCAGT TGATTTTGTG AAGTATGCTG AAGCTTTTGG AGCCAAAGGG GTTCGAGTTC	4200
ATTCACAAGC AGAATTTGCC GCAGCTTTAG AAGAAGGAAT GCAAACAGAA GGCCAGTGA	4260
TTATTGATGT TCCTGTGGAT TATTCAGATA ATCmAGAATT AGGTAAaCA TTGTTACCAG	4320
ATCAACTTTA TTAGGAGATG GAAAAATGAG CGAACAATAT GTTTATCAAC ATGGTACGCT	4380
AGGCGGTTTA ATGGAAAGTT TGATGGCAGG AACGGCAGAA ATTGGTACGT TACTCACGCA	4440
AGGTGATTTT GGAATTGGGA CATTAGAAGG CTCAAATGGC GAAATTATTT TATTGGACGG	4500
TACATTGTAT CATGCAAATC AACTGGCGA AATTACTATT CTAGAAGGCG AAGAATTAAC	4560
GCCATATGCC GCAGTTACTC GTTTTCAAGA AGATGGCGCA TTCCCTGTAT CAACGGAAAC	4620
CGATGAAAT ATTAAAGCAC AAATTTTAGA AAAAATTAGT CCTAATTTTT TTGCGGCAAT	4680
TAAAATTAGC GGTCTTTTTG CGAAGATGCA TGTCCGTGTG GCACCTAAAC AAGAAAAACC	4740
GTATCCACCA TTTGTAGAAG CAGCACGCAA TCAACCTGAA TTTACAGCGG AGAATATTCA	4800
GGGAACAGTT GTAGGGTTCT TTACACCTAA ATTATTTTCAT GGTGCCTCTG CCGCAGGgTT	4860
TCATTTGCAC TTTATCAGTG AGGATCACCA ATTTGGTGGG CACATTCTTG ATTTTGGCAT	4920
TAAACAAGGG ACTGTTTCGT GGATGGAAAC AGCAGAATTG CGACAGCATT TTCCAGTTCA	4980
TGATGCTGAT TATCGGAATA AAGAAATTGA TATTGCAAAA GCTTTGTCCG CAATTGAAGA	5040
AGCGGAATAA CTGAATAACA AGGCACATT CTCCACGscC ATTCAAGTTG ACTAACTAGG	5100
TCAACTTGAA TGGGCTTATT GCTGTTCGTC GCTTTTTGTT GTTATTTTTT CAAAAAAAT	5160
AGTATCCTTT TTATAATAAG CAAATAACAA TTGGAAAGGA TGGCGTAATG GACACACTTT	5220
ATAGAAGTTG GCAATTGTCT GGTGGCTCT ATCATGATAT TTTTGTATC ATTGTAGCAA	5280
TTATTTTTAT AGTAATTAGT GGTATTTTGG TTATTCCTT AATAAGAAGG CGGAGTACAA	5340
GGAGACTGGT TCCCTATGCG TTAATTTTAT TGGTTTATCT TGCAGTGGA CATTTTGCTG	5400
GTTAATCTT TTTCGGAATG TTTCGTTTCTG TAACGATTGA AGAGAAAAGT GCGACGTTTT	5460
ATTCTGAAAA AACAAAAGGC TTAACCTCAA TCGAGCGAAT GATTATTCCA AATGGACGAA	5520
CGAATGGCAT TAGCACCAGT AACTCATTGT TCCCAAGTGG ATTTChGTTA ATAGTCAGAC	5580
GGGCCGAACG AATGTGGnAG TAAACGTTTA AGTTGGCCGT GATTAATTTA ATGGnCCAAC	5640

AGATCCATAG GGGGTTTTGA nAA

5663

## (2) INFORMATION FOR SEQ ID NO: 53:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1799 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

AAACAGCATA GACAGCAATT GAGTTATGGC CmCgGTTtGA AGcwTwaAGG rAACgGCCaT	60
CAGArGAAaT ACGrATCGsA GcmCCmCCmT TAAACsCATC ATAATCAGCA GGTAAAGTAG	120
AAACTTTTTG TAAtTCaGTG AAAGaACCAT CTGTTyCATC ATATGAAAGA ACTGAAACAG	180
TGCTATCTAA TTCACCAAAT AAATAAGCCA ATGATTTTGT TGGGTGGAAC ACTAAATGAC	240
GTGGACCACT CCCTGGCTCA GTCACGAATT GAGCGATTTC AGAAAGCTTT CCATCTAAGC	300
TAACATTATA CGTATACACA CGATCGGTTT CTAGATCACA GACGACTAAG CGTTGATCAG	360
GCGTCAAATC TGTGTAATGA ACATGGGCAT TGTTTTGGTT TTCATGAGGA CCAGTTGGCT	420
CTGTATGAAT CACTTTGTCA GCATCTTCCA AAGAACCATC TTCTAAGATA CGATACACGT	480
GAACGATACC TTGGTGATAA TTTGCTGCaT AGACAAGTTG TCgGTtyTCG TCaACAGCTA	540
cATAAcATGG CGGCGCACCT TCCTCaGTTA CTTTAtTCaA TAAGTTAAAA tGGTCCATTT	600
CTGGTTGATA AGCGGCAGTG CCACCTTCAC CGTCAACACT AGTGACGCTA AAAAGTTCGC	660
CCTTTTGA CT CAGAGCAAGA TAGGTAGGAC TCGTTTCTTC TAACAAAAGT GTCGCTTCTG	720
TTAAACGACC TTGTGTCGTA TCTAATGCAA TTTTATAAAT GCCTTGCTG GCTCTTCGAG	780
TATAGGTACC TAGTAAaATT TGTTCaACA TCACAATTCC TCCAATTTGT TTCAACACAA	840
TCACTTTTTA TGAAAGTGCT TTTATATAAT GTAGTATACC aTAAGAAGTC TAAGAACAAT	900
AAAACTTTCT ATCTGAGATG TTCgCaTTTT AAAGtGGTTT GtGCtATTAT GaTgaGaGAC	960
tGTGaTGACa GTCaTTTTTT gaTTTTTAGA GGGGTGACGA ACATGTACGA CaATACgGAT	1020
AATAACGGgA CCGTTCGTTG TTCATTTTGT GGGAAACaC AAGAAGAAGT AAAAAAaATT	1080
GTGCTGGTC CGGGTGCTA TATTTGTAAT GAATGTATTG ATCTATGTAA AGAGATTATT	1140
GATGAAGAAT TTTATGATGA AGCAGTCCGA GAATTAACGG ATGTGCCAAA ACCCAAGAA	1200
ATTTTAAATG TTTTAAATGA ATATGTTATT GGACAAGAAC GTGCAAAACG CACACTTTCT	1260
GTTGCTGTTT ACAATCACTA TAAACGTGTG AATCAATCAG AAACAGCTGC AACTCAAGAT	1320
GACGTTGAAT TGCAAAAAAG TAACATTGT TTAATTGGAC CAACAGGTTT AGGTAAaACT	1380
TTCTTGCTC AAACGTTAGC AAAAACATTA AATGTACCTT TTGCAATCGC TGATGCGACC	1440
AGTTTAAACAG AAGCAGGTTA TGTAGGGGAA GATGTTGAAA aCATCCTTTT AAAATTACTA	1500
CAATCTGCTG ACTATAATGT TGAACGTGCG GAAAAGGGCA TTATCTATAT TGATGAGATT	1560



457

GATAAAATTG CCCGTAAGAG TGAAAATGTT TCGATTACTC GAGATGTTTC aGGTGAGGGT	1620
GTTCAACAAG CCTTACTAAA AATTCTTGAA GGAACAGTTG cAAGTGTACC ACCACAAGGG	1680
GGCCGCAAAC ATCCACATCA AGAATTTATC CAAATTGACA CAACGAATGT TcTCTTCCAT	1740
CGTTGGTGGG AGCATTTTGA TGGGAATTGG AAACCATGGT TAAAAACCC GCCTTAGGC	1799

(2) INFORMATION FOR SEQ ID NO: 54:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11266 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

CGCCAAGTTC GTCAATAATC GGTCAGATTG TTTTCAATCA GGAAATcAcG TTTTcGCCGC	60
TTCTTTCAGg TgAAAAATTTT CGCCCAGCGT TCATAGCCTG CGCCnTTTtC CGGCGGCAAG	120
TTTCGCCTGA ATGTCAATCA GGAGATTGAC CTTTCTGCCG GTTCGCTTCG GCGCTGTTTT	180
TCCTGACGTC AAACGTCTGC GCTGGCTGGT AATTCTGGCG CGCAAGGCGG TCTTGGTATA	240
ATCGTCGCCC AGCCGGTTCA CCCTTGTAGG GCGATCCCAT CCGGGCAGAC GGAATCCGGG	300
ATTTTtGCCG TCCACAACCT CATATCCCTC AGCCCGCATC TTTGCCAGAA ATTCCTCATA	360
ATGATTGCAC TCTGGCAGCA GCCGGTCAAT GGCTTGCTGT AACTGCTTTT TGTGgCTGAC	420
GCCGTACTGC GCCGCCGCCA TTTCTTTTtG GGGCCTGGCT TTTTCTTCG GATTCTCAAT	480
CACCGAAAGG CCGTGGGCGC GGCACAGTTC ATCATTCATG CGTCGCAGGA TAAAAGCGGA	540
GTTTTTCACG TTCTGGAATTT TTGCGTCACA GTTCAGGTTG GTGCTGTAA ATTCAATGTG	600
GGTATGGATA TGCCGCTTAT CCGTGTGGGT GGCCACCACA AACTGATGCT GCCCGCCGGT	660
AAATTTcATC GCCAATTcAC AGCCGATTTT GTGCGCCTCC TCCGGGCTGA CTTCATCGGG	720
TTTGAATGAC TGGATAATGC GGTACATGAT AATGTCATGC TTTTCTGAmC TGCTGCGCCC	780
GGTAAGCTGG GCGTACAGTT TTTTGTGAT TTCAAATTCT TCGGCGGCAG TTCCCGGGCT	840
GCACATATAG GAGGAAACCA GTTCGCCGCC GTCTGTCTTT TCTTCTTTCT GGTcATAGTC	900
GTGGCGTTCT GCCATTGACT GCTGACGGGT GCGGTTTTTc ATAATCTCTC GTGGTAAAAT	960
ATTAGAGATA GCCCAATGGT TCCTCCTTTC AGAAAAAGG CTATCCATAG CCTAAATAGT	1020
AGCATGGGTA GCCTACTTTT GCAAGCCATA TTTTGATGC TTGATACnAA AGGGGGCAG	1080
GATACAGCTA TGGATACGAT TTCAGCGATT AAGACCCGGA TACTCCGTTT ATGTGCGGAG	1140
CGCAACATTA CCATCAACAA ACTGGCTACG ATGTCGGCGC TGCCGCCATC CTCTGTCAAA	1200
AACATCTTGT ATGGAAAGAG TACCGATCCT AAAATTTCaA CCATCAAGAA AATATGTGAC	1260
GGTTTGGATA TGACGTTGAT AGATTTTTTc GATGCGCCCG AATTTGAAAA TTTAGArCCG	1320
GAAATTAAAT AGCGTTACCG CGCCGCTCC CGCTGCCGCT GTGCCGTCCG TACCGCTTTA	1380
ACGTGGGTGC AGAGAAACTC GTTATAGTCC TTGCCCTGCT TCGGCGGGTT GGGGTAAACC	1440

AGCTTCACCC GCCTGGAAAG TTCCGCATCG TTCTGGATTT CCTGGGTGAG CCGTTCCATG 1500  
CCCTCCAAGC CGGCCTTGTC ATTGTCGAGG CACAGACTGA CTTGGGTAAC GTGGGGGTGG 1560  
TCGTGAAGAA ACTGTATCAG GCGCGGGGG CCGGTTCCGC CCAGGGATAA ATAATGGCTG 1620  
TCCCGTGGTT CGCCGCCGGA CAGCTTTACT AATGTGGCGA GGGAGAGCGC GTCAATAGGG 1680  
CTTTCCGCCA CCGCCAGCCG GGGGCAGTCT GCCTTGCCCTG CCGGCAGACA GAAGCTGTAC 1740  
CGCTTGTCGC TTCCCTCCAC ATCCATGCGG AAGCCGTCCC GCGTCCCCCG CAGCTGGCG 1800  
GCGCGGGCGT TCCCTGCTGT GTCGCGCCCA ACAAAGACGC AGTTCTGGTA TTTGCAGCTC 1860  
TCGTATAAAA TCCCGGCTTC CATGCACACG CCAAGCAGTT CCGGGTCAAT GCCGCGATCC 1920  
TGCAAATAGC CCACAACAGC GGATGCACAT CGGCTGCCTG GGGGCAGTCG GAAGGGCTTT 1980  
GGCTGTTTTT CTTTATGGG CGGCGCTTTT TCGGGCGGTG GCGCACAGTA GCCGCACAGG 2040  
GTTTCCACCG CGCCTACAAA GTCCATGCCG CGCACCTTGA TCAGATAATC AAGGGCCGTG 2100  
CGCCCACCGA TGCCCCGGCT GTTCCAGCAC CATTTCCTGT TGGAGATTTT TAAGCTGTCA 2160  
TGGGTCTGG TGCAGTATTC GCGGGGGCCG CTTTTTTTCA GTTCCTGCGG CTCATAGGCC 2220  
TGCAAATAAG AGAGCAGATC CCATTCCTTT GCTTTTGCAA TCTGCTCTTT GGTACGCCT 2280  
GCCAAGTAAT CACCTCCAAA TATGAAAAA GGGCGGCGGG AGCATCTTTC TTTTGCCCCC 2340  
GCCGCAGTTG TTTACAGTGc TTCAAATTC TTCAGCAGGT CGTTCACGCC CTGCCAGATT 2400  
TCAGAGTAGC CCTGCCGCAA GTCCTCCACA TCCTGCTCAT AGAGGTTGTG GGTGCTGTTG 2460  
CACsGGCGGG CGATCTGGTT GACGTTGCTG GAAACGGAAC GCAGCAATTT TACCAGTTCC 2520  
AGCACCGCGC CCATGTCAAG CTGCACGATG TAGCCGTCCA CCGCCATCTT GCGTAGGTAA 2580  
GCCCCAGgT TGGAGGTTCC CAGCAATTCC ATTTTGCGGT CAATCGCCGC CTTTTCTTCC 2640  
GGGGAACTT TGGTGAACAG CCCACCGTG CGCCCTCCT TTATCCTCAC AGCACCGCCT 2700  
CCTTTCGCGG GGTGCGCTCC ATGCCGAAT GTACCGGCGT GGGCGGTTCA TAGTTACGGA 2760  
TTGCCTGACG GATAGAGGGT TTATAGCCAG CCGCCACAGC GGGCATCGGC ACAACCGGCG 2820  
CGGGTTTTCC GCGCTCCTCC AAATCAGGGC TTTCTTCAAC CTCCTGCTCA TCCTCCTTTT 2880  
TGTCGGCCT GGACAGTTCC CGGTAAAGCT GGTTTAAGCG GCGGATTTT TCCCGCAGTT 2940  
CTTCTCTTT TGGGAATGGC CGGTGCGCTT CTTCTTTGGC GTTTTCCAGT TCGCTTTCCA 3000  
GTGTGGCAAG CCGTGCTTTT TGTGCGTCCA ACCGCTCCGT CAGGCTTTCC AGCGCATTGT 3060  
TAATACGGGT GACGTTGCCC ACCGCATCGT CAGACAGTTC GCGGAATAG GTCAGGTTCT 3120  
GCTTCATGGT GACGCGGAAT TTGCTGCCGT CACACCTTAG CTGCATGGGA AAGCCGCGAA 3180  
ACTCGCCCAA ATCCACGGTT TTTTCCGGGT CGTCCATCAG CATAACGCT TTAATCACCG 3240  
cTTGCCCCGC CGCCTTGCGC TCCGTGTAGG TCTGCCCCAT GACAGTCATG GAAAACGCAT 3300  
CGTCCTTGAC CGGGTGGGCC TGCACGGTTG GCAAATCTGC CGTCAGCGCT TCGATTAAAC 3360  
GCTTGGTTTC CGCGATTTTC TGCGGGTAGT ATTTACGCGC TTTGTCCTGC ATTTCTACT 3420

GTTGGGAGGT GTGGTTGGCT TTGAGCATTT TCAGCTTGGA AACCTGCACA TCCAAATCCA	3480
TTTTTTCGCG GATACGGTCG TCGCCGGTGG CAAGGGCCTT GACCTCAGCA TAGGAAAGGG	3540
CAGTGGCGTC CACATCCTCG GCAGAGCGGG CAGGGCTTTT AGAGGTCATG ATCTGACCGA	3600
TGAATTTCTG TTTGTTTTCC ACCAGCCCCC AGTTGTAAGC GTCGAATGTA CCCTTGGTGA	3660
CATACTTGAA CATTTTCACC GACGCATTCA TGTTGCCCTG GCGCAGAGAT CGGCCCCGCC	3720
GCTGTTCCAA ATCGGCAGGA CGCCAGGGGC AATCCAGGTC ATGGGACGCG ACAATGCGGT	3780
TCTGCACATT CGTGCCTGCG CCCATTTTTG CCGTGGAGCC GATTAAGACA CGCACCTGCC	3840
CCCCCGCAC CTTGGCGAAC AGTTCTGCCT TTTGGGTTTC CGTGTGGCG TCGTGATAA	3900
AGGCGATTTC CGCTTCGGGG ATGCCTTTGG CAATCAGCTT TTCCTTGATG TCATCATAGA	3960
CATTGAATGA GCCGTCCCC TTGGGTGTGC TGAGGTCGCA AAACACTAAT TGTGTTCTTA	4020
AAATCTCCGT GCTGGCCTGC CATTCTGCA CGATATTTTT CACGCAGGCG TTGACCTTGC	4080
TGTCCGGGTC GTCGGGCAGG AGCGGATTTT GCAGCCGCTG ATCCAGCGCC AgCTTGCGCC	4140
CGTCCGAAGT AATCCGCAGC ATATTGTCGA TATGGGGGTC AACCTGCCCC CCGCGCACAC	4200
TTTCCGCACG CTCGCCAAGT TCCTCCACCA TTTCTTTCTG GAATCCGCTG GGTTCCTGG	4260
TGACGGTGAT ATATTCCGCC TTTGGCACGG GGAGATTGAG CATATCGGCG GTCTGGATGT	4320
CGGCGGCTTC TTTCCAAATA GAAATAAGTT CCGGCAGGTT ATAGAATTTT GCAAACCGGG	4380
TCTTGAGCG GAAGCCGCTG CCCTCCGGCT TTAATTCCAT CGCCGTGACC TTTTCGCCAA	4440
AGTCTGCCGC CCATGCGTCG AAATGGCGGT GGCCGTCTG CTCCAGGGTG TCAAACCTGGA	4500
GGTAGCGCAT CATGGTATAC AATTCTACCA TTGAGTTACT CACCGGGGTT CCGGTAGCGA	4560
ACACGATGCC GCGCCCGCCG GTGATTTCTG CCATGTAGCG GCACTTGGCA AACATATCGC	4620
TGGATTTCTG CGCGTCCGTC TGGGCAATCC CCGCCACATT CCGCATTTTT GTGTGTAAAA	4680
ACAAGTTCTT AAAGCCGTGC GCCTCGTCCA CAAACAGACG GTCAACGCCC AATTCTCTAA	4740
AGGTAACAAC ACTGTCTTTT TTCTTTCCGT CCGCCAGCTT TTGCAGCTTG GCTTCCAGGT	4800
TTTCTTTGT TTTCTCCATC TGCTTGATGG TGTAGCGCTC GCCGTCTCC GCCTTTGCTT	4860
CGGCAATGGC GGTGACGATT TCATCAATCT GCCCCTTAAT GACGGCTTTC TGCCGTTCCG	4920
GGGAGAGGGG GATTTTTTCA AACTGTGAGT GTCCGATAAT TATCGCGTCA TAGTCGCCGG	4980
TGGCGATCCG GGCAGCAAT TTCTTGCGGC GGGCAGGCTC AAAGTCCCGT TTCGTGGCAA	5040
CCAGCACCTT TGCGCCGGGG TACAGGCGCA AAAAATCGCC GCCCATTTGT TCCGTGAGGT	5100
GGTTGGGAAC GACGAATAGG GATTTCTGGT TCAGACCCAG CCGCTTGCCC TCCATTGCGG	5160
CGGCAATCAT TTCAAAGGTC TTGCCCGCGC CGACACAGTG GGCAAGCAGC GAGTTTTTCC	5220
CATAGAGCAT CCGCGCCACG GCGTTCTCTT GGTGGGTACG CAGGGAGATT TCCGGGTTC	5280
TCCCAGCAAA ACGGATATGG GAACCGTCAT ATTCGCGTGG ACGGATGCAG TTGAATTTCT	5340
CGTTGTACCG CCTGCAAAGG GTTTCCCGAC GTTCCGGGTC TTGAAAATC CAGTCCTTGA	5400

ACGTTTCGCA CATGGCTTCC TGCTTCTGCT GGGCAATGGC GGTCTGCTTT TCGTTGAGGA	5460
CGCGCACCTC CTTTCCGTCC ACTTCTGTCT TATCAAATAT CCGTACATCC CGCTGGTTTA	5520
AGGTCTGCTC AAAGATTTC AAGCGTTGA TGCCTTGGT GCCGTAAGTA GCATAGACGC	5580
GCACGTTATC CCGGTTGTCT GTGCTTTTTC CCTGCACGTT CCATTGCGCC GTGGTAGGGG	5640
AATATTGCAG CTTGATTTTT CTGTCTGTGA AATACACAGG TGTATGCAGC AGTTCAAACA	5700
TGAATTGCTG GTAATACTCC GGGTCAATCC AAGACGCGCC GATACGGACG CTGATTTTCAG	5760
ATGCCGTCAG GTCTTTGCGC TGCACCTGTT CCAGCTTTTC GGCATTGATG GCAAACCTGCG	5820
GGTTTTCCTC AGCGGCGGCG CGGGCAATCG CTAAGTTGGC CCGGACGTTG CCGGAAAGGT	5880
ATTCGTCCGC CGTCTGCCAG CCCTGATGCC AGTGTGCGCC GCCGTCTCC AGGTCAAAGG	5940
GGCCGGTGGC GGGGTTTTTG AAGATGATGC CTTTTAAGTC CTCCACAATC TGCGGGATTT	6000
TTTCAGGGCC GCCCATGAGG GATGCCATAT AGCCCAAATC CACGCAGGCT TTTTCACCGA	6060
TAGATACCGC CAGCGCTCC ACAGCCGTGT CCACGCTGGT AACGGCGCGG TGGGTCTGGA	6120
TGGTGCGCCT GGTGAACATA TCCGCTTTAC GCTCCAGCTT CCCCTCCTCG TTGATGATTT	6180
CAAGAGAGCA GAGCAGCGGA TAGGAGGAAT CCTGCTCAAA GGCCAGCTTG TTGCCAGTGC	6240
TGTTAAGAAG CCCGTACTTT GCGCTGAACG CATCATAGAG CCGGTTTAAT TTTGCCTGCT	6300
CGGACTGGAT TTCTGTGTCG TCTGCGCCGG CAAGCTGTAA GTCGATCAGG CGGCGGGCGC	6360
TGTCCCGGAT GCCGATCATG CCCCTGACGC GCTCCGTGGG GGTTTTGCCA AGGACGGCCT	6420
GGGTCATGCG GGAGTTTTCC CGGAAGTACA GCTTATCATT GGATAGGGTG TAACTGAAAT	6480
TGCGTACAGT GGGGTCGGCG GGGATGCTTT CCAGTTCCTC GCCGTCTTGC CCGGAATCCA	6540
CTTCTAACAG TTCCCGGTCA GGCGGTGCGA TATGCCGAT AGCTTCTGCA AGCTGCCCCG	6600
AAAGGTCTGC GCCGGGGATG GGCTGGCAGG CGGTTTCAGT TTCATTGCCG TACATACTGC	6660
GTCCCATTGT CATTGTCCCC AGCACCATTT CCGGCTGGTC TAAAAAGTAG CGGTTTCAGG	6720
GGATGCCGTC CTCTGTCTGC CCGACCTCCA CCCATTCCGG GAGTTTTTCC GGGATGCGGT	6780
CGCGCTTCTG GAAAAAGAGG ATGTCCGTTG TGAATTCGGT TCCGGCGTTG GCCTTGAACG	6840
CATTGCTGGG CAGGCGGACT GCACCCAATA AATCGGCTTT TTGTGCCAGC GCCTCCCGCG	6900
CCTTGCTGTT TTTCTTGTCC ATTGTGCCCT TGGTGGTGAT AAAGGCCACA ATGCCGCCGG	6960
GACGCACCTT GTCCAGCGTT TTGGTGATGA AATAGTCGTG AATCATCAGG TTTTGACGGT	7020
CGTACCGCTT GTCATGCACC TGATAGCTGC CAAAGGGGAC GTTGCCCAACC GCAAGGTCAA	7080
AGAAATCGTC CGGGTGGTCG GTGCGCTCAA ATCCGTCTAC CGTTATATCC GCCTTTTGGT	7140
ATAGCTGTCT GGCGATGCGC CCGGTCAGGT CGTCCAGTTC CACACCATAT AACTTTGCTG	7200
CCGCCAGTTT TTCGGGCAGC ATCCCGAAAA AGTTCCCGAT GCCCATAGAC GGCTCCAAAA	7260
CCGTCCCCGG CGTAAAGTCC ATCTGTTCCA CCGCATTGTA GATTGCCTGG ATAACGGTGG	7320
GGCTGGTGTA GTGGGCGTTG AGAACCCTGC TTTGTGCAGA CTGGTACTCC CCAGGGGTGA	7380

GCAGTTCTTT CAGTTCGGCG TATTCCTTTG CCCATTTTGG GTCGTTTGGC TCAAAGGCGT	7440
GGGAAATGCC GCCCCAGCCC ACATAACGGG ATAAAATGGC CTGTTCTTCT GGTGTGGCAA	7500
GGCGGTTTTC GGCTTCAATC TGTTTCAGGG TGCGGATGGC TTCCACGTTG TACTGGTACT	7560
TGGTCTTTTC GCCGCCTGCG CCTAAATTTT CATCGGTGAT GCGGAAGTTG TGGCGTTCCG	7620
GCTCCAGACG GAGCGGCTGT AAAATAATCT CGCTGGGGAA TACCCGGCTG TCCACTTCCT	7680
CCTGGGCGCG GCGGCGGGTC TGTCTTTGCG CCGGGGTTTC CGTAATCTGC CCGCCGTCAA	7740
TCTCCACCAG TTCCGGCGGT GCTTCCTGTT CTGGAGAAAC GTCCTTCTCT GGAAAAATTT	7800
CCGGCTCTGT TGGCTCCACG CCCTCTCTGG CACCTGCCAT TTCTGACCAC CTGATCCGCT	7860
GGTTCAGTTC GGCTTTTAAC CGTTCTTCAA TATCCGGTTC CACGGAAACC AGCATATCAT	7920
AGTCAAGGAA AGCAAGTTCC ATATCAATCA GTTCCCGCAG GCTGCCATAC TGCCGTGTTT	7980
CCACGCATTT CCCGTCTACG GACTGGCTGA TGGAGAAATC CAGCAGGTTT ACGTCCACCT	8040
GGACTTCTAA ACCGCCGTCC TCGGTCGTGC CATAGGCAAT CCCGATATGC TCCAAATCGG	8100
AAAAATCCAC ATCATCGCTG CCATATTCGG CGTAGGCAAA CTGTTCAATT AGCTGTTTTG	8160
CCGTATCTAA ATCGGAGGTT TCGGGGCGTT CCGGCTCCTG CATGGGTTC GCTGTAAAAAT	8220
GCTCCTGCTG GACTTCCTCT ACAAAGTCC GCACAAAGGG GATGCTTTCA CTGCGGAAAA	8280
TGGGATACCA GCCCCGAGT TCCATATCTT GCAGGCTGAC GGTTCGGCGG TTATAGTCCA	8340
CGCTGTCAAT CTTCATCAGC CGCCCGTCAA TGGTCAGTTC CATGCCGACA GGGATATAAT	8400
CCGCCGCGCT GCGTTCCTTG ATGATTTTAT AGGGGGCATC GCCGCCGCGT TCCGTGTCAG	8460
GCCGCGCCAG CACCACGCTT TTTCCATGTT CCAGCAGCTT TTTTAAGACG GACTGCCAGC	8520
CATTGGCTCC GGTGATGGAC ACCATCCCCA GGTGCGGAAT ATCGCGGGTG ATGGTTTTCG	8580
TTCCCAGCGC AGGGGCGGCT TCCTCTGCGT CTGTGCCGTA AAACAGCATG GTTTTCCCTA	8640
CCCGGACGCC TACCAGTTTG TCAGGTATT GGGCTTTCAG GTTCAGGTAT TCCTCCACGT	8700
TGGGCGTAAG GTTCAGACCG TCTGCGGCGG CTGTTTTTGG GGCTTCCTGG GTCTGCGGCG	8760
AACCTGTGCC GGTGGGTTCT GCGCCCGCCT TTTGCGCGGG GGATGGTACA ATATTGCTGT	8820
CCCCGGTTTC CGGCTCTGGA TTGACCGCAG CGGGCGCGTC CTGCACCTTT TCCGGCTCTG	8880
GCTCCGAGTT CCCaGGCTGC ACCCGTTCCG GCTCCTGGGG CGGGGTGGAA AATACTGGCG	8940
TGTCCTGGGT GGCAGAAATG GACATTTTCA TGGCAGCGGC AACCTTTTGA ATTTTTTCCC	9000
TGATTTCTCT CGGCATATCC GCGTCGTAAA AAGTCACGGT ACGGTCAGGG GCGATATGGG	9060
CAACTGTTTT GTAGTCGCCG TGTTCTTCCT CCAGACGGTT CCAGACGGTC AGGCCGTTCC	9120
CCATATGGCC ATAGCCTAAA TCATAGCGGT CAGCGGGGCT TTCGGCCATT ATGGGCGCAG	9180
CCCCTTCTGG TATGTTCCGC GTTTCCTGCG GTGGTTCCGG CTTTGATGCG ATAGGGACGG	9240
CCGTTTCAGA TGCGCCGGTT TCTGCCTGCG AAAGTTCCGG CTCCGGTTCA GGAGTTGCGG	9300
CTGGCTGTTC TGTTTCGGAC GAAAGCCCGA ACATATCTAA AGTAAGCTGC CCGTCTGCAT	9360

CCTCCCGTGG GGCAGGCTGT TTCTTGCGGG AAAGGGCTTT TGCCGCAGCA GCAAGGCCGG	9420
TTGCAGGTTT AGCCGTTTGT CTGGCGGAAT CCTGCCTTTC CTTCTGTTTC TGTTTCTGCT	9480
TTTTCTGCCG TTCCTCCTGC AAGACGGCAT CAGGCAGAGG GGTGAATAAG GAGTATTCGC	9540
CCCGCAGgTA TGCGCCCAT TTTTGAAGCG TCGGCTCCAT CAGGGAATAG GCGCGCTCAT	9600
CCGGGGATAC CATTGCCAGG GCTTTCACCA TTTTTCATA GTGCGCCGCG CATTGTTCCG	9660
GGTCATCTAA CAGCGGGCGG AATACCTTTT TTGCCGCGTC AACATCCCAG TCGTGGGGGA	9720
TTGTCTTGCT GGCATCGTTG GGGTCGTAGT ATTGGAAGGT ATATAGACTC CGAGCAAGCT	9780
GCATTTTTTC ATATTCCGGC AGATATGCCT TTTCCcTGCG GTTCAGATAC CGATCCGTTT	9840
TAATCAGTCC GGCAATCCGC TTTTGTACCT GCTTCCAGTT CAGCGTAACC GTATCATAAT	9900
TGCCACCGGA AAAATCGTCG GCGCGGGACA GCTTGATGCC TTTGCCGTCG TGCCATTTCAT	9960
CATAGCCAAT GTAGCCATGC CCGCCAATGC CGTATTCGTT TTTCAGAAAT TCGATGCACT	10020
CTTTGGCATC ATGCCCTGC ATAAAGTAGG CGTAAATACG GAATTTCCCC TCGGAAACGC	10080
CAGAGCCGCC AGTGAACATC CGGGTAATCT CGTCCTCGGT GACAAAAGCT CCACGGGCAG	10140
GCGCAAAGCC CTCCACCCCA TGAAATGCCG TCACCGGCTT ATCCATATCG GTCAGGCGGG	10200
TAAACAGTTC GGAGGGCGCG GGCCGCTGGC GGAAGCGCAG CAGATTCGGG TTCTCCTGAT	10260
GGGAGATGGC AAACCTGCCG ACTTCCTGTG CAATTTTGGC TCGTTCTGCC GGGTCTTTCA	10320
GGAGCGCGGC AATTTTAGCC GTGCTGTCCG GGAAACCACC GTACAGCGCG TCAATGGTGG	10380
GCAGCAAGCC CTTTTCTTTT GCTTCCTCGC TGAAATTCTG CCGCAGAAAC CACAGTTTTT	10440
CCGACAATTC CCGAAATTCA TTGTCACGGG CCGCGTCGAT TTTGTCCTGG CTGGCAAAC	10500
GGCCGCTGTT CAAAAGCTGC CGGATCTGTG CCGCCGCTG CTTCCAGGTG ACAAGGGAGC	10560
CGCCCCGGC AGAGCGCCCC GGCGCGATGT GGATACCGCT ATGGCTGAAC CACAGCGAAT	10620
ACTCATGCCC GGCAATTTTA AGCCCTTGC CGCCGCTGCC GTATTCTTTT TGCAGAAAAT	10680
CGGCGGCAGA CTGCTCCGTA GGGTCTTTCT GGAAAAAGC GACGATACGC AGGATGCTGT	10740
CCGGCTCGTT GCCGCCGCTT GTGAGGGCGC GGCCAATGAC CGCATCCGTG AcCTTGCCGG	10800
AGGAAAcGGC GGrGCGCAGT TCCTGCTCGG TTTTCTGTGC CTGAGCAATA GTTTCAATCT	10860
GtGTTCCACA GTGGGAATA AAGAAACTG TGTAATGCC GGCTTTTCAG AcGCAGAAAm	10920
GcGGACTGTT CGCCCGCCGC TTCCTGCTGG TTCTCCTGAT TTAATTGTAA ACGATCTCGT	10980
TCAGTACGGT TTCCTCCGCC TGGGCGAGAA ACAGGTTGAC TGCTGCACC CATCCCATCT	11040
GGTCGCGGGC TTTTCATCCCC TCGTmCACGC CGTTCTGTTT CTTAATTCC GCCACGATCC	11100
GGTCTGCCTG TTCCTGCGCh TGCCGGTCGA TCTCCGCGAG GTGGGGTTCC AGCCGCCCGG	11160
TCAGCAGCAT CGAGGTGTAT GTCCCTTGT GGTGTTCTG AAGAAAGGTC TGCCGCAGCC	11220
TGCCGTACTT GCCAGCGGC ATTtCCTCCA TCTCGTCTC CCCGTC	11266

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7528 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

AATTTTACTG TAGTAAAGAA GTACGTGAAT TTGATTTTGG ATATTGTATT AGGGATTTTA	60
GGAGTTTCT TATTAGTATT TATGGCTAGT TATTTAGTTG ACCTCGCTAC ATATCTCATG	120
AAGCCAATGA CGCCGGAIAA TTTTTCAGTT ATCATGCAAG AAATTATTTT ATTCTTTATG	180
TTATTTGAGT TTATTATGAT GATTCTTCGC TATATTCAAG AAGGACACCA TATTCCGATc	240
GTTATTgATT TTAATTgTTa ACCGCAATTT TACGTCAATT AATGGkTGCG CATGGCGATG	300
GCyTGCAGAC TTWAtkGtTa wCTTTAwCAA TTCTTTTACT AGTGGTTGTG CTATTTATTC	360
TCGGGTAAAA TGGCAGCAAA TTTTAtACGC TAGGAAAAGC GAAAGTAGaA gAAGaTCACG	420
AGCaAGATTT TCayCATTAa TCgAACaTAA TCaAAGACCG TCCAACGTGA GAATTTTTCT	480
CTAACGTGGG GCGGTCTTTT TAATGCATAG TTAAATAGTT AACTCAATCA GTTCAATTGT	540
TGGTAAGACA CCAAAGCGTG CGCTGATTTT AGTGGTTCCG ATTCCTGGGT GAACAATTAG	600
CGGCGTTTCA TCTTCTAAAT GATAAAGCCC ATGTGTATAT TTTTAGCAA GCTGATTTTT	660
GTAAGTGAAA AAAGGCAAGC GAATCTGTCC GCCGTGACTA TGTCCTGCTA AAATGAGGTC	720
AATGTTTTTG TTGACTGCTG AGTCTGCGAC ATCTGGCTCA TGGGTTAGTA GCAGAGTGGG	780
CGTTCCAGTA GGGCGGTTGA CTAAAGCTGC AGCTAAGTCG GTTTGTCCAT ATAACGAATC	840
TTCTAAACCA GCGACGTACA AGGAGATACG GTGTGAGACA GGCAGTAATT TGCCCGTgT	900
TTTTaGTACG GTAAAGTCGG CAGCTTCTAA GATTGtTTCG TAAAAATTGG TATCTAAGCC	960
GCTATACTCG TGGTTGCCAT AAATAGCGAA TTTGCCATAA GGGGCGTTAA TTTTACGGAA	1020
AAGTTGACTC ACTTCATCTT GAATTTTCGAT ACCGAAACGG GCATAATCAT CAAATAAGTC	1080
TCCTGTAAAA ACGACTATGT CTGGTTGTAA TGTATTGTGTT TTTTCAACAA TCTGTTTTAA	1140
GGCCTTCACA GAATAATTCT TTTTAAATG AATATCAGAG AGTTGAACAA TCCGTAAGTT	1200
TTCCCTGATT CTGTTACTTt AATTTGATTT TTGTGtATGA CTAAGCATTT AGGTTCAATA	1260
AAAAAACTAT AGCTTAAAT GCCGATAATA ATGAAAAATA GTGCCAGAAA TAAATAAACG	1320
TTCATAATTT CCTCCTTGAA GTCGCTAAAG TAGTTTACTA CAAAAGTGTT AAGCAAAAAA	1380
TAAGGAAAGG TTAAATTTTC AGTAACTTT CTCGATGAAG TTTTTGGTG ATTCATGCTT	1440
TTCTGAAGTA AATTTGCTAT AATGAAAAAT GAGAAAAAAG AAAGAAGGTC GAACCAATGA	1500
TGGCAAGAGT CGAAAAGCTA CGGGAGTTAA TGAAAGAAAA TAATTTACAA GGATTTTTAG	1560
TGACAAGTCC GTATAATTTA CGTTATTTAA CCAATTTTAC TGGGACmACA GGCTTGGCAA	1620
TGATTACATT AGATAAAGCT TTTTTGTAA CGGATTTTCG TTATaCGGAA CAAGCTGCGG	1680

AACAAGCAAa CTGGTTTTAC TATcGTTAAA AACACAGGAC ATaTTTTTTGa CGAAGTAGCA	1740
GATTTAGCTG AACGTTTGCA ATTAGATAAT TTAGCTTTTG AAGAAACGCA AGTTAGTTTC	1800
GCTGATTATA GCTTATTGGA AGAAATTTTA CCATGTGAAT TAGTTCCTGT AATGGGTTTG	1860
ATTGAAGAAT TAAGAGAAGT GAAAGATGAA GAAGAAGTAG CGATTATTGA AAAAGCGTGT	1920
GCCATCGCGG ATCAAGGATT TGCCTTTGTC TTAGAGATGA TTAAACCTGG AATGACAGAG	1980
ATTGAAGTGG CTAATCAATT AGATTCTTC ATGCGTTCaA AAGGAGCCAG TGGTGTTCCT	2040
TTTGAGACAA TTGTGGCCAG CGGTCTACGT TCAGCGATGC CACATGGCGT TGCTAGCCaT	2100
AAAGTCATCG AAAAAGGTGA GTTAATCACT TTAGATTTCG GTTGTTATTA TGAAGGCTAT	2160
GTTTCTGATA TGACACGGAC ATTCGCTATT GGTAGTATTC AGCCAAaACT AAAAGAAATT	2220
TATGACATTG TTTTAGAAGC ACAATTAAAA GTGTTAGCAG AAGCTAAACC TGGGTTGACG	2280
GGCATTCAAT TAGATGCGAT TGCACGGGAT CATATTGCTT CCTATGGCTA TGGCGATGCG	2340
TTTGGTCACA GCACTGGTCA TGGCATTGGG TTAGAAATTC ATGAAGGACC AAATGTTTCA	2400
TTCCGAGCAG ATAAACAGTT TGTACCAGGG AATGTGATTA CGGACGAACC TGAATTTAT	2460
TTGCCAGGAA TTGGTGGCGT GCGTATTGAG GACGATTTAT TAATTACTGC TGAAGGTAAT	2520
CGTGTATTAA CACATGCGCC AAAAGAATTA ATCATTTTAT AAAGGTTGGG AGAACATGGA	2580
TATCCAAGAG TATCAGCGGT TTATTAGTGC GTTTTATAAA AAACGTAaCT GGTATCAATA	2640
TGATCCGTTT ATTCGTTCCA ACTTTTTAAC AGAAGAAGTC GGCGAAGTTT GTCGGGCCAT	2700
TCGGACAATT GAAATTGGCC GTGATCGATC GGATGAAACA GTTTTGGCGG ATGAAGAGAA	2760
TCTGGAAaAT TTAACGGAAG AGTTAGGCGA TGTGTTGGAT AATTTATTTA TTTTAGCCGA	2820
TAAATATGAC ATCTCTTTAG AAATGATTAT GCAAAAGCAT CGGACGAaAC TAATGGAACG	2880
TTATCCTGAA GAATAAAGAT GTAAAAATAA GAAAGCCAAA ATTACTTGGT CTTTCTTATT	2940
TTTACGTCTA AAGGCTTGAA TTAGAAATTT ATGACAAGAA ATTATCTGTT CTTTtagTCA	3000
GAACTTAGCT TTTCTGGTAG CGAAATAGGC TAATTAATGA TAAACTAAGA ACGAGTATTG	3060
ATAAGCAAAA TTTGGACGTT CTTTCACAAC TAACTGAAGA AACCAGTGTt GTTTTAATTT	3120
TtAAGAATCA TCCGTTGCTT AAAAGAGAAT CCTTAAGAGA CTTAGTGATT CATCAaATAA	3180
ATCGTTTATT TAACGATTTT GTGCGTAATA GATAAGGAGG ACAACCAaAT GTCAGAAGAA	3240
AAAAATCTTG TAATCAATGC AAACGATTCT TTAGGAGAAA TCGTGATCGC TCCTGAAGTC	3300
ATTGAAGTAA TTATCGGAAT CGCTGCTTCA AAAGTAGAGG GCGTTTATGG GATGCGCGGC	3360
ACGTTTGCCA ATAATGTGAA TGAATTATTA GGCCGTGCAG CACACGGCAa AGGTGTATAT	3420
TTACGAACAG AAGAAGAAGG CTTAAAAGTC GATATCTATT GTTACTTGAA CTACGGTGTT	3480
TCTGTACCAa AAGTTGCGTT AGAAATGCAa GACCGTGTGA AACAACAAGT CTTATTTATG	3540
ACAGATATTG ATTTAGTTGA AGTGAATATT CATGTTGTAG CGGTAGTACC TGAAAAATC	3600
CCTCAACCAG ACTTAAATGA ATTATTCCCA GAAGATGAAG ATGGAGAAGA GAATGAGTAA	3660



AACAGAATTA ACCCGTCACG AGATTCGTGA AAAAGCGCTC CAAGCCCTTT TTCCTTTAGA	3720
TTTTAATGCT GATTTAACCA AACAAAGATGC CATTGATTAT GCATTGGCTT ATGATAATCG	3780
GGAAATCGTC AGTGAAGATG GCGAAGACCT AGTCCCTACG TATTTAGATT TACTTGTCGG	3840
CGGTGTGTGT AGCCGCAAAG CAGAACTAGA CGAAGTGATT ACGAACCATT TAGGTAATAA	3900
CTGGTCTATG CAACGCTTAG CAAAAATAGA TATTGTTATT TTACGTTTAG CTATTTTTGA	3960
AATGCTATAC GTTTCAGATG TGCCAAACAT TGTGGCATT AATGAAGCCG TTGAGTTAAG	4020
TAAAAAATAC AGCGACGACC GTTCACGTAA ATTTGTTAAT GGTGTGTTAT CAAATGTAAT	4080
GAAAGAAATT GATTCAGAGG CGTAAGCTCT GAATCTTTTT TTAATATCTA ACTGGTGTTA	4140
AGATACTGGG TTATATGCTA AAATAAATAC TATTAAAGAA GTAAAGAAAC AATAGAGGGA	4200
GCGAAGGAAT GAGTACAGTA ATTAACGGTC GTGAATTAGC AGATCAAATG CAAGCAGAGA	4260
TTCAAAAAGA TGTAGAAAAG ATGACACAAC AAGGCATCCA ACCAGGATTA GTTGTTTTTAT	4320
TAGTTGGGGA AAATCCTGCC AGCCAAACGT ATGTGAGAAA TAAAGAACGT GCAGCAGCCA	4380
AAATTGGCAT TCTGTCAAAG GTCGAAAAAC TGCCAGAAAC TATTTAGAA GAAGAATTAT	4440
TGGCTGAGAT TGACAAATAT AATCAAGATT CACGCTTTCA TGGCATTCTT GTACAACTAC	4500
CTTTGCCAAA ACATATTGAT GAAGAAAAGA TTTTATTAGC GATTGATCCC AAAAAAGATG	4560
TAGATGGGTT CCATCCAATG AATTTAGGCC GTTTGTTTGT AGGCAAACCT GAAATGATTC	4620
CTTGCACGCC GTATGGAATT ATGAAGATGT TTGAAGCTTA TGATATTGAT TTAACAGGTA	4680
AACGTGCGGT GGTATTGGT CGAAGCAATA TTGTTGGCAA GCCAATGGCA CAATTATTGT	4740
TAATGAAAA TGCGACCGTG ACGATTGCCC ATTCTAAAC TGAACATTTA GCAGAAGTTG	4800
CAAAGAAGC GGATATTTTA GTTGTAGCAA TTGGACGCGG GCATTTTGTC ACCAAAGAGT	4860
TCGTGAAACC AGGTGCGGTA GTGATTGATG TGGGCATGAA TCGCAATCAA GAAGGAAAAC	4920
TGATTGGTGA TGTAGCATTC GATGAAGTTT CTGAAATTGC AAGCTATATC ACGCCGGTTC	4980
CCAAAGGTGT TGGCCCAATG ACGATTACCA TGTTAATGTA TCAAACAGTC GAGGCAGCAA	5040
AAAAACAAAA GTAAGGTGAA AAAATGACGC AACAGTATTT AACAGTAACA GCCTTAACGA	5100
aGTATTTAAA ACGTAAATTT GATGCAGACC CTTACTTAGG ACGTGTCTAT TTAACAGGTG	5160
AAATTTGAA CTTTCGCTTC CGAGCGAACG CGCATCAGTA CTTAGTTTA AAAGATGATC	5220
ATGCGAAGAT TTCTGCAATT ATGTTTAAAT CAGCCTTTCA AAAATTGAAG TTTCAACCAA	5280
AAGAAGGCAT GAAAGTCATG GTGGTGGGGC GTATTTGCTT TTATGAAAAC AGCGGCTCCT	5340
ATCAAATTTA TATTGAACAT ATGGAACCAG ATGGTGTAGG TGCACTTTAT CAAGCATTGG	5400
CAGAATTAAG AGAAAACTA GGAAAAGAAG GACTTTTTGA AGGGCCTAAA CAACAGTTGC	5460
CTCGTTATCC CAAACGAATC GCTGTGGTTA CAAGTCCTAG TGGTGCTGTT ATTCGGGATA	5520
TCATTACGAC AGTCAAAAGA CGGTATCCGA TTGCACAACT GGTCTTTTTT CCAACCTTGG	5580
TCCAAGGCGA GCAGGCGGCG GATGACATTG TGCGGAACAT TCAGCGAGCC GATGCGCAAA	5640

466

GCGATTTTGA TACGATGATC ATCGGCCGTG GCGGTGGTTC AATCGAAGAT TTATGGCCAT	5700
TTAATGAAGA AAAAGTGGCT CGAGCCATTC ATGCTGCCAC CACACCGATT ATTTATCGG	5760
TTGGTCACGA AACAGATGTG ACAATTGCCG ATATGGTAGC AGATGTTCTG GCAGCAACGC	5820
CAACGGCAGC AGCGGAACATA GCGGTACCAG TTTTAAATGA AGAACTATTA AGAATTAGTG	5880
AACGGCGTAG TCGTTTAGAA CAAAGCTTTT TATATCTCTT ACAACACGA ACAGAACGTT	5940
TCCAACGACT ACAAATTCT TATGTGTTA AGCAACCTGA GCGACTTTAT GAAGGTCAAA	6000
CGATTAAATT AGATCGAATG ACGCAACGTC TTTTCCAAGC AATGACACAA ATTCATCACC	6060
AAAAACAACG TCAAGCCCAA GGAATTATCG CTCAATTGCA ACAACAGACA CCAAGGGAC	6120
AATTAAGAGA AAGTCAGCAA CAATTGGCAT TTTTACAACG AAATTGCAA ACACAAATGA	6180
CACAATTGTT CCTAAACAAA CAAAGCAAT TCACGTCTGC TGTTCACAA CTAGACTTAC	6240
TCAGCCCACT CAAATTATG GGTGAGGAT ATAGTTATAC TACAAAGAG GACCGTGTG	6300
TTAAACTGT TACCGAGTTA CAGCCAGCAG ATCAATTAA CATTCAATTAT GCTGACGGTA	6360
CAGTTCAAGC AACCGTCGAA ACAATAACAG CGAAAAAGA GGAGTTTAA CATGCCAGCA	6420
AAAGAAAAA CGTTGAAGA ATCATTAAAT GCCTTAGAAG AATCGTTCA GCGCTTAGAA	6480
CGAGGAGATG TTCCTTTGGA AGAAGCCTTA GCTGCCTTTC AAGAGGSCAT GGCATTAGT	6540
AAACAATGTC AAGATACATT GGAAAAAGCG GAAAAACTT TAACGAATAT GATGACAGAA	6600
AACAATGAAG AATTGTTTT TGAAGAGAGC GAGGAAGCGT AATGACGAAT TTTAGTCAAC	6660
AGCATTTACC GTTGGTTGAA AAGTCATGG TCGATTTTAT CGCAGATAT ACTGAAAATG	6720
AGCGTTTGAA AGAAGCAATG CTCTACTCAA TACACGCAGG TGGCAACGA TTACGCCCGC	6780
TATTAGTGTT AACAACAGTG GCCGCTTTTC AAAAAAGAGT GGAAACCAA GACTATCAG	6840
TGGCTGCCTC TTTAGAGATG ATTCATACGT ATTCATTAA TCATGATGAT TTACCAGCAA	6900
TGGACGATGA TGATTTACGT CGTGGCAAC CAACCAATCA TAAAGTGTTT GGTGAAGCGA	6960
CTGCCATTTT AGCAGGGGAC GGCTTATTA CAGGTGCATT TCAGTTGCTT TCTTTGAGTC	7020
AATTAGGCTT AAGTGAAAAA GTTTTACTGA TCAACAACCT GGCAAGGCA GCGGGGAATC	7080
AAGGCATGGT TTCCGGCCAA ATGGGTGATA TTGAAGGAGA AAAAGTCAGC TTGACTTTAG	7140
AAGAGTTAGC GGCTGTTTAT GAAAAGAAAA CCGGAGCGCT AATTGAATTC GCCTTAATTG	7200
CTGGGGGCGT ATTAGCCAAT CAAACAGAAG AAGTCATTGG TTTACTGACA CAGTTTGCGC	7260
ATCACTATGG CTGGCTTTT CAAATTCGTG ATGACTTATT AGATGCCACA AGTACAGAAG	7320
CAGATTTAGG TAAAAAGTA GGCCGAGATG AAGCGTTGAA TAAAGTACG TATCCAGCTC	7380
TTCTAGGGAT TGCTGGCGCG AAGATGCGC TACACATCA ATTAGCAGAC AAGCTGTGAC	7440
CGTCTCCGGG AGCTGCATGT GTCAGAGGTT TTCACCGTCA TCACCGAAC GCGCGAGACG	7500
AAAGGGCCTC GTGATACGCC TATTTTAA	7528

(2) INFORMATION FOR SEQ ID NO: 56:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6682 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

```

TTTTTTTGCA ACGTTTCCAT GAAAGAATGC GGAAGATAT CCCTAAAAAT TGATCTTCTT      60
ATTTCTCTAT TCAAAATGAT TCTAGCCTAA CCTTTTTTAC TATTTTCATGA TATGATGAAG      120
CAAAGAATAT TTTAGGAGGT TTTTACATGG CTGCAAAAGA AGCAAGTTTt GATGTAnTTT      180
CAGAGGTCAA TATGGAGGAA GTCAAAAACG CAATTCAAAT TGCCTTGAAA GAATTGAAAA      240
ATCGTTTTGA CTTTAAAGGG TCAATTGCTG ATATTAAATT GGAAAACGAT AAATTAGTAG      300
TCGTAGCTGA AGATGATTAT AAAGTCGAGC AAGTGAAGGA TATTTTATTT GGaAAATTAG      360
TCAAACGCAA TGTTCGATT AAAAATATTC ATTTTTCAGA AAGTGAAAAA GCCTTAGGCG      420
GTACTGCTCG CCAATATGGT GATTTAATCA GTGGTATTGA TAAAGAAAAT GCAAAAAAAA      480
TCAACACAGC TATCAAAAC TCTGGGATTA AAGTTAAGTC ACAAATCCAA GAAGACAAAA      540
TTCGGGTAAC TGGAAAAAGT CGGGATGATT TACAAAAAGT CATGGCTCTT TTACGAGAAT      600
TAGATTTACC AATGGcTCTT GAGTTTAATA ATTATCGCTA AAAAAATCCT TCAACTGTTA      660
ATGAACAGTt GAAGGaTTTT TTTGTTTTtGG TACCGGCCAT TCTTGATTAT TCAATAAAaC      720
GATTGGTCGG CGTTGGATTT TTCCTGAAAA AGCGGCAATT GCTTTGAAAA CCTTTGgCAT      780
AAAATAAAAC GCATTtAATT CTTCTAAATC CAGTTCTTCT TCTGCCGTAA ACGAGACACA      840
AACCTCTTGT TCTGAAAGCA TAATTTCAAT TTTCTTTAAA AATTGAAATT GCCGCATTCC      900
TGGCACGAAA TAACCAGCCT TTGCAACACG AGAGGTTGTA ATATCTAATT TATCACTAAT      960
GATTAGCGCT AGTGTCATTA ACTCTTCCGA ATCAAAGCCG TTACTATGAT TTTCTATAGA      1020
AGATAAAATT TCTTCTTGAT AAGGCAAACA AATCTTTTGc TTACGAAAAt ATGCTTCTGC      1080
AAAttGTTTC CCACGAAGTG CATGTCCTGC TTTTCCTTGT AGCGCACCCA CATCATGGAG      1140
CAGTGCCGCA ATTTTTGCCG CCTCTAAATA CTCCTCTGAA ACAGCAAGTT GTTTCAAAAT      1200
CATTTCGTC ATGGCTACGA CATTGGTTAC ATGAAACCAA TTATGGTATG CCCACCCAGC      1260
TTCTTCATCT TCCATTGAC CAATTGCATC ATAAATTTGT CGAATTTcAG TATCTTGATG      1320
AATGCGTTCA AACAGCTTCA TTTATTTTCT CCTTAAATT AGTTACTGTT CCAAAAATTC      1380
TTTTAACTCT GCAGATTTT TTTGCTGTT TAAACCAACA ATTAAACGAA TCCGCGCTTT      1440
CGGTCCGTTT AAACCACGGG CAAAAACGAC CCCCATTTTT TTCAAGCCAA CGCCACCACC      1500
AGCATAATCA TAGATATCTT CAGCAATGCC ATTGGAACAA CGAGAACTA AAACGACTGG      1560
AATCCCCTTG TCTAGCATTT TTTGTAATGC TGGGAGCGTT TCTGGCGGTA AATTACCCGC      1620
GCCTAGCGCT TCAATTACTA ACCCATCCGT TTCTGGGGTG TTTAATAATT CAAACATCCG      1680
TTCACCCATC CCCGCATAGG CcTTCcTACG TGAACATTTT CTTTCACAGA aGAAACATCG      1740

```

CAAACCTCTT	GCGGTAAGAC	CTCTTTCGCA	AAAAAAGCAC	GTTCTTTCGC	AATGGTTCCT	1800
ATTGGTCCAA	AAGTTGGCGT	CCGAAACGTC	GCTACGTTTCG	TTGTATGTGT	TTTTGTCACA	1860
TAACGCGCCG	TATGAATTTC	ATCATTTCATC	ACGACTAAAA	CACCTTTATC	GTACGATTTCG	1920
TCTGAACAAG	CAGTCCAAAT	AGCACTGATG	AAATTATACA	AGCCATCACT	ACCGATTTCG	1980
TTACTTGACC	GCATTGCTCC	TGTTAAAACA	ATGGGGATTT	TTTTCTCTAA	GGTTATATCT	2040
AAAAAATAAG	CAGTTTCTTC	TAATGTATCT	GTTCCATGCG	TAATAACCAC	GCCATCAATT	2100
GGTTCAGAAT	AAGCTTTTTG	AATCCGTTCT	TTTAATTGAA	ACATCCGTTT	TAAGGTCATG	2160
TGTGGTGAGG	GAATATTAAG	AATATCTTCA	ACAACATAAT	GAACCTTGCC	aGAAAAATAAT	2220
GCTTCTTGTT	CCATCAACGG	ATTTTTTTCA	TTCGGGGCAA	CACTTCCTCC	TGCTACTTTC	2280
GACATGGAAA	TGGTCCCACC	TGTATGAAGG	ACTAAAATGG	TCTTCATTTG	TTTCCTCCTT	2340
TGTTTTCTTT	ATTTTTTAAC	CAAAATTGTT	TCACTTGTTG	AATGAGTTGT	TCATTGGTGT	2400
TTGCCTTTTCG	ATGCGCCAAC	CAATGCGGCG	GAAAGAAACG	CTGGACACTT	GGTAATTCAA	2460
ATCGTTTCGTC	TAATAACAAC	ACAAAGCCTT	TATCTTGCGC	ATCTCGAATG	ACACGTCCAG	2520
CCGCTTGCAA	GACTTTATTC	ATGCCaGGAA	TTTGATACGC	AAATTGAAAC	CCTTGCCCTC	2580
GCTCTTCTTG	ATAAAAAGTG	CGAATCAGTT	CTTGCTCTGG	GTTCAATTTG	GGTAGCCCGA	2640
CACTAACAAT	TATCGTACCA	ATCAATCGGG	TACCTTTCAA	GTCAATTCCT	TCTGAAAATA	2700
TACCACCTAA	AACACAAAAA	CCGATGAAAG	TTTCTTGAGG	ATTCTCTTTA	AATTCAGCTA	2760
AAAATGCTTC	TCGCTCTTTT	TCGTTTAAGT	CCGTTCCCTG	AATTTGAGTT	TTTACTTGCG	2820
GGTAACGTTG	GCTAAATACT	TGATAGACAT	CATCCATATA	GGCGTAGGAA	GGAAAGAAAA	2880
TTAAATAATT	ACCTGTTTTT	GTTTCACTTA	ACTTCCCTAA	AAGTTCACTA	ATCTGTGGAT	2940
AACTTTTTTC	ACGGTTTTTA	TAGGTTGTTT	GAAGATGACT	ACCAATCATC	AACAATTGAT	3000
TTTCTTCCGG	AAAAGGATAC	GGAATCCGAT	AACGTAGGCT	TTCTTCACCT	CCACCAAGCA	3060
CTTCTTGATA	ATAATTGAGA	GGTGTTAAAC	TTGCCGAAAA	TAAGATACTC	GCTTTCCTT	3120
TATCTAATCG	TTGTGACAAT	AAATAAGCCG	GATCGACACA	AAATTGTTTC	ACAACGCAGT	3180
CATAGTTTCG	CAAAGAAATA	GACGTATAAA	AGCCGTTGTC	GTAATACTCG	CTGATTGCAA	3240
GATAACTTAA	ACAATCAAAA	TAAAACTGA	GAATCGACGC	CAATTCCTCA	ATGTGTTGAT	3300
TTTCTGGCAA	CCATTCTGCA	ATTACTTCTG	TTAATCGATG	TATCTGTTTA	ACTAATGAAT	3360
CAATTGGCGC	GCGTTGGGCA	AAAAAATCCG	TGTCTTTTTT	TTACACAGATT	GCTTCAATAT	3420
CATCAAACGT	AGTTAAGACT	TTTTGACAAG	CTCTTTTCAA	TTGAACAGAT	TCTTTTTGCA	3480
ATCTTTTTTG	AATTgCAAAA	CCGCATTCCCT	AGAAATAGCG	GCCGAATACA	TCTCCCGTGA	3540
ACGATTAACC	AAATTATGGG	TTTCATCCAC	TAAAAAATA	TTCTCTTCTG	cACTTTGGGG	3600
TTCTTCAAAG	AAGCGCCGTA	AATAAACCGT	TGGATCAAAT	AGATAATTGT	AATCGCCAAT	3660
AATTAAATCA	CACCAAAGAC	TTGCATCTAG	CGACAACTCG	AATGGACACA	ATTGGTGCTT	3720

TTGTGCATAC CGTTCAATAA CCTCTCTCGT TAAATGATTT TCATGTTGCA ATAAATCCCA	3780
CAGTCCCTCA TTTAAACGAT CATAATAGCC ATCAGCAAAC GGGCAATGTT CTGGTGTACA	3840
ATTTCTGTTCA GTTAAAAAAC AAATTTTATC TTTTGCTGTT AACGTAATAC TCTTAGTTTG	3900
CGCTTGTTTT TCGGACAGCG CTTTGACGGC ATCTTCAGCC ACTTGCCGCG TGATGGTCTT	3960
AGCTGTTAA TAAAAATCC GTTCACCTTC TTCTTCACCA ATTGCTTTCA TTGCTGAAA	4020
AAGAGTGGAC ATTGTCTTGC CTGTTCCAGT TGGTGCTTCC ACGAAGAGCC GCTGTTTACT	4080
AAGAATTGTT TTATAGACGG CTACCGCCAA TTCTCGTTGC CCTTGCGAT ACTCCCCGTA	4140
AGGAAACTA AGATTTTTTA ATGATTTATT TCTCAAATA CGCCACTCTT CTTTAAAAAT	4200
TACCCACTGT TCATAGCGCT TCATCAAATC AGTAAAAAC TCCGCTAATG CAGCTTCTGT	4260
AAAATGGCGT TCTTGTCGTG TGATTTGTTT TGTGGTGGTC TGATAATAAG TGAGTTGAAG	4320
CGTTATTTCA GAAAGTTGTT CTTGTTGGCA ATACATGTGA CCATAACACA TCGCTTGATA	4380
CCAAAACAAT TCGACCTTTT CTGGCGGCAA TTCTTCAAAC GCTGGTTCGG ATGTTTTAAT	4440
TTCATCAATG ACCACTTGCC CAGTTTCTTT CACAAAAATG CCGTCTGCAC GACCTTCTAC	4500
ACAATAATTA TGGCCATTTA AAGGCACCTC AATTTTTAAA GATACTTCCT TTTGATAAAG	4560
GCTCCCCGCT TCCTTTTGCA ACATTCGATG AAGGCGACTA CCTTCTTG TGATATGTTT	4620
ACTCACATGG CGAGAATCGA TACTTCCCCG CCTTAAAAATA AATTCGACCA CTTGACGCAC	4680
CGCAATTTTT CTCGTTGATT TCACGTAATC CCACCTATCT GCCTCCCTAT TATAGAACAT	4740
ACATTCGCTA AACAAAAGCA AAAAAATCAA AACGAACACA ATCATGCGTT CATTCTGGAT	4800
TTCTCTTTAG ATTATCAACA ATTAAAAACA ACTTGTTTCT TCTAACTGTT TGGTTTCTTT	4860
TCCATTTGCT TTTGCCACGA CCAATTGCGC CATTGTGTTA AACGAATTAT TTAAACCAAT	4920
GCCGCCCTTT TTGGGCATCT TCACAATCGG TACTTGCTGA TAATAGTGAC TAACTGCTGG	4980
ATTTTCTTCT GCCATCGCTG CAATTGCAGG AATGCCTGCA GCATTAAACT TACAAGCTAG	5040
ACGAGCAGCC ATTTCTCAA AATTAGGATA ATGCATGGCT GGTCCGCAA GAACCGCATC	5100
GGCATGAAAC TTTTTTGCA AGCCAATAAA TTTCTTTGCT ACCTCTTCCT CGTTTTCCAA	5160
ATAATATTGG TCACCACAAT AAAGTGTCGC AACAATTTC GCACCTTGCT TCTGTAAAAA	5220
TGGTAGCAAT GTTACCCCTG GACCTAAGGG AGCTTTTTTA CCACTTGGTT TCAGCTGAGC	5280
TTGTTTCATCA CTGCCAAATC CTGCTAAACT GTGGTTTAAG ACTAAAAGGA TTTTTTTCAT	5340
GCCTATTTCC TCCTTATTTA TAAATCATCG AAGTCAAAGG AATCAAAATC AATTTCTTCA	5400
CTCTCTtCTT CTTGTGCTTT CCATTCTTCA CGTAAATATT GGTATCCAT AACTCGAATA	5460
AAAGGATAAT AAATGAACAT TCCTAGTACC AATAAGAGAG CTGCAGAAT TGATCCTTGC	5520
CAACCGCTGA CTAATAATCC TGAGATAATT GCGGGGTCG TCCACGGCAA TTGAATACCA	5580
TTCGTCAATG GAACTAAGCC CCAATCCATA GCAAAATAGG CAATAATAAT ATTAATTGCG	5640
GGTACTAAGA CAAATGGAAT TAAAAGGATT GGATTTAAAA CAATTGGTAA CCCAAAAATG	5700

ATTGGTTCGT TAATCCCAAA AACACCAGGA ATTAGTGAAA GTTGCGAAAG CTTTTTAATT	5760
CGTTGaCTTT TACAGAATAA GACCATCACA ATGATTAATG AGAGTGTGGA aCCGGCGCCA	5820
CCAAATGTTG cAAcATATCT TGAAATTGTC CAGTGATGAT GTTGGGAATT TCATGACCGT	5880
CTTTAAATGC TTGTAAATTT TCAACAGACA ATGCCCCGAA AACAGGATTA AAGACTGCGC	5940
CAACCACACT ACCTCCGTTA ATGCCAAaGA ACCAAAAGAA GTGTAAAAAT AAATAAGCAA	6000
TGCTCATTGC ACCTAAGGTA TTACCAAGAC TCAATAAGGG CGTTTGTAGA AATTCAAAAA	6060
TAACATCTTG TAAATTACCA TAAGGAGTTA AAGCAGCCAG CCAATTGACT AAAAACATGA	6120
GAGTAACAAC TAAAAAaCTC GGAATCAGGG CCTCAAAAGA CTTAACCCT GTTGGTGGA	6180
CTCCTTCTGG CATTTTTATT GTCCACCCTT TACGAACAAT GGCCACAAAT AAACGAGTCG	6240
CAATTAATGC GACAATCATT CCTAAAAACA AGCCACTGGA ACCCATCCAT TTTAAAGGTA	6300
AACTAGTAAC TTCATAAACC TTTGAaCTAT TTTCCGGTGT AAATAATGTT TGATAAGGCG	6360
TTACAATAAA AAATGAAACC AATGCGATAA TTCCTGCTTG CATTCATCC ACATTGAATT	6420
GTTTTCTTAA ACTATAGGCw ATACCGACAA CGGCCAGAAT AGTCATTACA TCAAaACTAG	6480
CACTAGCTGG AACAGACATC ATTGTCGCCC AGTCTTTACC AAGCAAATGA GACATCCAAG	6540
TATCCCATG TGGCAAAGGA AAATTAGCAA TTAATAAGAA GACCGAACCA ACAATCAACA	6600
AGGGCGTTGA TAAAGGAAG CCATCACGAA TCGCTACTAA ATACTTATTk TCTCTATCA	6660
CTTTTGCCAA AGGGGTCAAT AC	6682

## (2) INFORMATION FOR SEQ ID NO: 57:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18724 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

AACTCCTAGA ACCAAGTAGG TACTGTTTAA CATCTATTTG CTTTGCAAAT AGTGTTCCTC	60
AGCAATAAGA AGGGATTCTT GAAAATTGTT CTATAATTTT TGGAGACCAA GTAGGTATTG	120
TTCATCATCG AATCACTACG TTCTTCGTGA TTTCAACAA TTTCCGCTTA TTTCCGAAGA	180
CCTTCAACTC TTAGAGGTTT AGCTCTTAGA AGTTGTTGAG AACGAAGCAA CGCGTAGTGT	240
TGCTTCTGTT CCCGCCGTTT ATCAGTTTTT GAGCGTGAAG TAAAAATCCA AAGTGATTTT	300
TATCCACGC CCATTTTCAT TTATGACGTT GGCTTCGTGT AAATTGAGAA GTCTCGTCCT	360
CTCTAGTGGT TGATTGCTGT GTTCGTTGTA CGGCTTGCTT TTCTGGCTCG CTCTTTTTTG	420
CAACATCAGT CACAACGTTT GGCTTCTGAG AAGGGGGAAC AGTTGTTTTT TTGTTTAAAG	480
TTGGCGTAGT TCTTTTTGTT TCTTCTTTT GTTCTTTTTT ATTATTTTTA TCATTTCCAG	540
TGACTGTCAA ATAGATAATC AGAAAAATCA TTAGGATCGT TGCAACAATT AATAGAATAA	600

ACATGTTAAA TGAAATGAAC ATGCATGGCC CTCGTTTCTC GTTTTTCTt AGTTTACCAA	660
AATtCaaaa TTTCGTCACT CTTTTCTtCA GAAAAACGA ATTTCAAAGG GGTtAAAAGA	720
AACTTTTATA AAAAAGCCTC CTTTTTTGTG TGAAATGCA CAAATTTCT GGaATCGTTT	780
GCTGTTAAAA CCAGATACTG GTAAAATAAT TATTGTATAA GaATCaTTTT AAAAAACGAT	840
TTTTGTTACA AATAACGTAT TTGTTGAAAA GGcACTGTGA AGTAGTTAGA GACGACAGCC	900
aAGTGACTTT TAATGACAGC CGCCCAATTA ATTTCAATGA GCCTTTTCGT AGAAAACGAG	960
GAGGTTATTT TGCCTGAAG TGAGTATATT TTCGACATGC GTGGTGGATT TACTTTTTCC	1020
AAATGTAGGG CAAGCAATGG TAGAGTTTT GGAACGTTAT GGCTGTGAAA CGGTTTTACC	1080
AACAAGCCAA ACGTGTGTG GTCAGCCAAC GTATAATAGT GGCTATGTGA CAGAGAGCAC	1140
GACAACGCTT AAGAATCAA TTGATAGTTT TGACGGTGCC GATTATGTGG TGGGACCTGC	1200
GGGGTCTTGT GTCGGCATGA TGAAAGAATA CCACAAATTT TTAGCAGATG ATCCTATTTA	1260
TGGACCAAAA GCACAACGTT TAGCTGAGCG AACTTACGAG TTTAGTCAGT TTCTTTATCG	1320
TGTCTTAGGG GTCAAAGATG TTGGCGCAAC GTTACATGGG AAAGCCACTT ATCATCGTTC	1380
TTGTCATATG ACGCGAATTT TAAATGAACG AGAATCACCC TTTGTTTTGT TGGATCATGT	1440
GAAAGGATTA GAAATGATTC CTTTAGGACA TCTTGAAAAT TGTGCGGAT TTGGCGGAAC	1500
GTTTTCACTG AAAATGCCAG CTATTTCTGA ACAAATGGTG ACGGAAAAAA TGAACGATGT	1560
GATTGATACA GGCGCAGAAA TTTAATTAG TCGGATATG GGTGCCTGA TGAATATCGG	1620
GGGTAAATTC AATCGTGATG GGAAAAAGAT TAAGATTATG CATATTGCAG AAGTCCTTAA	1680
TCACGAAGTG GATGAAGCAC GCATGGACCA ACCACAAATC ATATCGGTAG GGTAAGGGGG	1740
AAGAACTGTG GGATTATCAA CGAGCAATAA GCCGTTAAGT GAACGAATTG AAGAAAGTAA	1800
AAAAGATGTG TTTATGCAA AGGCTGTGCG TAAAGCACAA GATGCCCAAT GGGAGAAACG	1860
AGAAGGTGCG CGGGAAGCTT TAGGCAATTG GCCACAATGG CGTGAACCTG GAGAACAAAT	1920
TCGCCAACAT ACCATTCACT ATCTGCCTGA CTATTTAGAA GAGTTCAGTG ATAACGTTGC	1980
CAAACGAGGT GGCAAAGTCT TTTTGTCTCA AACCGCCGAA GAGGCCAATG AATATGTGAA	2040
ACAAGTGGTT CTAGAAAAAA AGGCGAAAAA AATTGTGAAA TCTAAATCAA TGGTGACGAC	2100
CGAAGTCGAT ATTGATCCAA TGTTATTGGG ATTAGACGAT GTCAGTGTGA TGGAAACAGA	2160
CTTGGCGGAA TTCATTTTAC AAATGGATGA CTGGGATGAG CCGTCACATA TTGTTTTCCC	2220
AAGTATTCAT AAAAAATCGCG AACAGATTCG TCAGGTTTTT GCGAAAAAAT TAGGGTATCA	2280
AGGAGATAAT GACCCGGTAA ATTTGGCACG CTGTGCACGA GAAGTGATGC GCAAATTCCT	2340
TTTGGAAGCG GAGATTGGCA TTACAGGTTG TAATTCGCT ATTGCGGATA GCGGCTTAAT	2400
TAACCTAAAT ACAATGAAG GAAATGCTGA TTTGACCATC AGTATTCCTA AGACACAAAT	2460
CGTGTTAATG GGGATGGAAA GAATCGTACC AACAATGCGT GAAGCAGAGG TTTTAGATAA	2520
CCTTTTAGCC AGAAGTGCCG TTGGTCAAAA TTTAACAACG TATGTGACGT TTGCCGGGCA	2580

AAAAAATGCG GATGAATCCG ATGGTCCAGA AGAATTTAC GTGGTCATTT TGGATAACGG	2640
ACGTTCAAAA GCGTTAGGAA CAGCCTTTCA GCCAGTCTTA CAGTGTATTC GTTGTGGGTC	2700
TTGTCTGAAT GTTTGTCCGG TCTATCGCCA TATTGGAGGG CACGGTTATG GCTCTATTTA	2760
TCCAGGACCA ATTGGTGCTG TGCTTTCTCC AATTTTAGGT GGATATAAAC AGTTTGGGGA	2820
ATTGCCATTAT GCATCTAGTC TTTGTGGGGC ATGTACGGAA ACGTGTCCTG TGAAAATTCC	2880
ATTGCACGAA TTATTAATTG AACATCGTAA AGTGATGACC GATGATCTGA AAATGAAACA	2940
TGGATTTGAA GATTTCCAAA TCGGTATGGT AGGTAAAGCC ACTGGTTCAC CAGCAATGTT	3000
TAAAGCAGCC ATGAAAGTCG ATCATGCTGC AGCGGGCATT TTAAGTAAAC AAAAAGACAT	3060
TACTGTTGAA AACATGTACA ATCaTGGAGG CTACCTTGAT AAAGGACCAG GGTAGTAAA	3120
AGGTTGGACC GATGTTCTGT ATTTGCCAAG ACCACCAAAA TCTTCAGAAA ACTTCAGAAG	3180
TTGGTTCAAAA AAACATCAAG AAGGTGAAAA AAATGACTAA CGAAGCGATT CAAAAATCGAG	3240
AACCTTTTTT ACAAATCTA CGAGAAAAAT TAGACGTAGA AAAACAGCCT GTTTCGGCCC	3300
ATCCATTTGA GCCAGTTAAT CATTTACCAG AAGAACAGTT GGCAGATAAA ACACCAGCAG	3360
AGCTATTAAC GATTGTGAAA GAACGTGTGG AAACGATTCA CACGAACCTA GTGGAAACCA	3420
CACAAGAAAA TTTACTGACG ACGATTACAG AAATTGTTGC TGAGTTTGGT GGGGACAATC	3480
TTTTATTGCC AACAGATGCA CGTTTTGAAG CCTATGGATT AGCTGATTTA GCTAAGTCAT	3540
TAGAGGCTGT TTCTGTTAAG CAATGGCAAC CTGGAAGTGA GCAACGAGAA GCGAATATTC	3600
AAACAGCTGC TCaGGCaAAT ATTGCGATTG CCTTTGCTGa GTTTTTAcTA GCTGAATCaG	3660
GAACGATTGT GGTGGAGTCC aATGCGGGAC aAGGCAGAGC CtTACATTTT TTACCGAAGC	3720
ATTACATTC AATTATACCG TTTAGCAAAC TCGTTCCACG TTCGACTCAG CCAGCAGCTT	3780
TTTATACAGA AAAAATAGAA AAAGGCGAAA AAATCGGCTC AGCGATTAC TTTATTTAG	3840
GGCCTTCTAA TTCAGGAGAT ATTGAAATGC AACTGGTTGT CGGGTTGCAT GGTCTTTAG	3900
AAGTTTGTTA CGTGGTTGTG ATGGATCGCT AAAACGCTAG AAGCATAATC AATTCAGCCC	3960
TTTTTAGAAT TGATTATGCT TTTTTATGA GCAGGATCAA AAAAGTAAGA ATCAAATTTG	4020
AAAGAAATAC GTTTTTTTTC TGACCATTTT GTAAAAAAG CGTACATAAA AATTAATTGG	4080
GTGTTAAGCA CTATAGGATT GATTTACTTT GTGGTAGAAT ATGAGACATA AGATAATAAG	4140
GTTTTTGGAG GAATGAGTTT GAAATCAAAA GAATTAATTA AAACAGTCGT CTTTTTTGCC	4200
TGTTTAGCTT TGGGTCTGTT TTTACTGAGA CAATTTGTAT TTACGCCTGT CGTAGTGAGA	4260
GGTCATTCAA TGGATCCAAC GTTAGCAGAT GGTGAACGGG TAATTACGTT AAAAAACACA	4320
GAAATTAATC GTTTCGATAT TATTACTTTC CCAGCGCCAG ATGAACCAGA TAAAAATTAT	4380
ATTAAACGTG TGATTGGTTT ACCTGGAGAT ACAATTGCGT ACAAGGATGA TACGTTGTAC	4440
ATCAATGGAA AAGAAGTTGA CGAACCTAT TTAGATGAAT TAAAAAAGC CTTAACAGAT	4500
GGTCAACCTT TGACAGGCGA TTTTTCATTA AAAGAAAAAG TACCAGCAGA TAGCTACTTT	4560



GTTTTAGGTG ATAATCGACG GAATTCAAAA GACGGTCGTG TCATTGGTTT TATTCATAAA	4620
AAAGATATTT TGGGTGAAGT GAAATTTGTG ATGTGGCCAT TCTCACGGTT TGGTCCAATA	4680
CCAGAAGTGT CAAAACAATA AAAGAGACAA TGAGTAAAAG AGAAGGACGG AAATGTGTAA	4740
AAACAAACAC ATGTTTCCGT CCTTCTTTTG TGGTAAAATA AAGGCAGTGC AAAGGAAAAG	4800
GATGTGAAAC CATGAGTGTG CAGTTTATTA GAGGAACCGC CGTGGCAGAC TTAGAAGCGC	4860
CACTTATACA GGCGACAAAA CAATGGCTAG AGGAAGACGC ACAGCATGAA GTATTCTATT	4920
TAGTCCCGAA CCATATTAAG TTTGAGCAAG AAATTCAAGT CTTGCAAAAG TTACGTCAAT	4980
TACAGACAAC TACGTCTGAT TCAATAACCA GTACCCGTCT GCAGGTTTTT AGTTTTTATC	5040
GTTTGGCTTG GTACTATTTA CAACATACAC CTTTTACTC GGCAGATGTT TTATCTGATG	5100
CAGGTGCTGC GATGATTTTT CGCAAAATTT TAGTAGAAGC AGAAGAAGAG TTACAAATTT	5160
TTAGAGGTGA AATTAACAAG CCTGGCTTTA TTCaGCAAct ATTCAGCTT TATCaAGAAA	5220
TGCGCGAAGG CAATATTGAG ATTGCGGAGC TTTATCCATT TTTAGAAAA CAAACGGAGA	5280
ACCCTAAAGG ACAAGATTTA CAACTTAAGT TTCAAGATTT AACGTTAATT TTTACGCGAT	5340
TCCAATTGCA AATGAGTCAG TATGGCTATG AATCAGCGGA GATTATTCAA CATTTAAGCG	5400
AGTATTTACA AACGGTTGAT TTGTCGAATG TTCAATTCGT TATTTAGGC TATCAGCAGT	5460
TTACTGCCAG AGAATTAAAG TTGATTGAAG TGCTAATGGC GCAAgcAGGT TCTGTGAAAG	5520
TAGCCCTTTT ATTGGACAAA CAATATCCGC ATGACTTACC TGATCCGCGG TCGTTATTTT	5580
ATGAAGCGGG ACAACCTAT CATCAGCTTT ATCAACTAGC AAGACAGAAG CAAATCCCGA	5640
TTCTTTCAGA TTATGTGGAA AAGAAAGAGG TATTAATTAC AAATCCAGAT TTGCAAGGAT	5700
TAAACGATTA TTGGATTCAA TCGCAAGAAC ACCTGCCTCC ATTAAGCACA GCTGACTGGA	5760
GGGGCGATGG CTTATTTCTG TGGCGTGCTG AAAACGTTAA AGAAGAATTA ACACATGTTG	5820
CAACCGAAAT TCGCCGTTTA GTGGTGGAAG AAGGGTATCG GTACAAAGAA ATTCAGGTGT	5880
TGACTCGTGA TTTAGATTGT TATGAAAATT TATTAGAACC GATTTTTGCG GAACACGAGA	5940
TtCCTGTTTA TGTCGACCGA GATATGGCTA TGGATCGGCA TCCTTTAGTC GAATGGATTG	6000
AATCTTTATT TGCTATTCAT TCGTACAATT ATCGTTATCG CGATGTGCTC CGTTTTTTGA	6060
GAACGGAATT ATTTATGCCA ATGAATCAAC TCGCTACTTC TGAAGAAAGC TTGACTGATT	6120
GGTTAAACCA ACGGAACGCT TGGCGCAGAA AAGTTGATAT TACCGAAAAT GTAGTCCTGG	6180
CATACGGCTA TGAAGGTTAT TATTGGTCAC AAGAAAAGA TTGGGAATTT ATTCGGTATG	6240
ATTTTGAAGC GGAAGAGCAA GAAGATGTCG CTACGATGGA AGAAGAGTCA AATGCGATTG	6300
GTCAATCGCT TCAACGGCTG TTACCTAGTT ACTTTCAAGC AATGATAAGT GCCAAAAC TG	6360
GTTTGAAGC AGCGACCGTC TTCTATCACT TTTTACTGCA AAGTGGTGTG GCAACACAAT	6420
TAAAAATGTG GCGCTTACAA GCTATCGAAG CAGGTCAACT AGAAACAGCT CGTAACCATG	6480
AACAGACTTG GGATGCATTG ATGTCATTAC TAGATGAATA CGTCACTGTT TACGGAGAAA	6540

GTTCTGTTGA TTTTACAACG TTTCAAGAAA TATTTGTTAG TGGTTTGGAA GGTTTGCACT	6600
ATAGTAAAGT GCCGACAGCG ATTGACCAAG TACAAGTTCG CGCAATGGAT TTAACGAC	6660
CTGGCGCTGC CAAGGTTACT TTTGCTATTG GGATGACGGA AGAAATTTT CCACAAAAGA	6720
TTGAAAATAA AACACTATTG TCTGATGAAG AACGGCAGAC AATCAATGAT ACTTTAACGG	6780
AAAACCAATA TTTACGTGGG ACAACCGGAC GCAAAATTGC GCAAGAACCT TATGTTGCGT	6840
ATCTTGTCTT TTCTTCAGCG CGAGAACGGT TATATTTAAC CTATCCAAGT GTTAAAGATA	6900
CCGCTCAAGA AGTCAAACCG TCTCCTTATT TTAAAAATAT CCAAAAAGAT TTAAACTTGC	6960
CCGTTTTTGA AAAAAATGAA ACAACCATTT TCGATGATGA AACGACCACT TTAGCCCAT	7020
TTAGTACGTA TCGGACGTTA ATCGGTGAAT TAACGCGTTT GAAAAGACAA CGGAAAGAAA	7080
CACAAGAAGG ACTGTTGCCT TTCTGGTTAA ATATGGAAAA AGCGCTAATG AACCAATCAA	7140
TTGCGCCGCT AGCCAAGCAA GTTTTCGAAA GTTTAACGCA TCAAAATATT CCTGAAAAGA	7200
TTGATGAGGT CTTAGCAGAA CCTTTATATG GTAAAGATAT TTATACGTCC GTTTCTCGGA	7260
TGGAGAGTTT TTATCGCTGT CAGTATCAAT ACTTTTCTCG CTATGGCTTA CGTTTGAAAG	7320
AAAGAGATGT CTTTGGGCTT TCGCCAGCAG CTACAGGGGA ATTTTTTCAT GAAGCATTGG	7380
ATCAGTTCTT TAAATTATTG ATTATGAATC AACGGAACCT ATCTGAATTA ACCGATCAGG	7440
AAGTCAATTT ATTGGCGGAG GAAGTACTAA ATAGTATTTT AGGGGATGCG CGTTTTTCTG	7500
TATTAACGAC CTCTAGTCGC ATGAACATA TTCGCTATCA ATTGAGTCAA ACAATTAAAA	7560
AAGTCAGTTG GCGCTGAAA CGTCAAAGCC AACGAAGTGA CATGACGACT GTTCAAACGG	7620
AAGTGTATT TGGCCAAATT GCCGCTAAAA AAGGAATTTT TGGTTTGGAA TTACCATTAA	7680
AAAATCAAGG GAAGATTcAT GTTCGGGGAA AAATCGACCG GATTGATCAA TTGGTTACCC	7740
CCGAATCAAC CTATTTAGGT GTGATTGATT ATAAATCGAG TCACCGTAAA TTTAATATGA	7800
CAGAAGCGTA TTACGGCTTA GCCATGCAAA TGCTGACGTA CTTGGATGTG GCTTTAATGG	7860
ATGCTGTGCA ATTAGTTGGT CAAGAAGCTA AACCTGCTGG CTCACCTTAT CTGCATGTAC	7920
ACAATCCTAC ATTATCTTAT GAAGGCAAAG ACGACATTGA ACAACAAATG CTTAAAAAAT	7980
ATCAGTTTGA TGGTCTTTTG ATGAAAGACC CAGATTTATT GGATCATTTA GATACGAGTC	8040
TGCAAGCTAA ACAAAGTTCG TTGCTGTTTC CGATTGAAGA ATCCGCCAAA GAACAAATTA	8100
AACCTGGTCG TCGCCAAGAA GATAAATTTG TGACAGAACC AGAATTAGGC GCGTTGCTTT	8160
CCCATAACCG AAATAAATTT ATTGAAGCGG GGAATCAAAT TATTGGCGGC GAAGTTCAAC	8220
TTAATCCCGC TTATCAAGGC AAGGAACGGA TTGCTGTGCG CTATTGTCCG TTTCGAAGTG	8280
TCTGTGATTT TGATGTGATG CTAAAAGAAA ATAATTATCA TCGTATTGAG AATTTATCCA	8340
AAGAGGAAAT TATGGCGCGC TTGTAAACA AAGATGAGGA AGGAGCAACC GAAGATGAGT	8400
AAAACAATTC CACTACGTCC AGCCAATGAG CAATTCACCG ACAGTCAATG GCAAGCGGTG	8460
TTTGATGGCG ATGAAAATAT CTTAGTTTCT GCTTCGGCTG GCTCAGGAAA AACCACCGTT	8520

TTAGTCCGCC GTGTCATTGA AAAAGTCAAA AGCGGCGTCG ATATCGATCG TCTCTTGATT	8580
GTGACGTACA CTGAAGCCGC CGCCCGTGAG ATGAAAGAAC GTATCCAAGT GGCTTTACAA	8640
AAGGCAATGA ATGAAGAACA AGATCCAGAA AGAAGACGGC ATTTTCTCG TCAAATTGCT	8700
CTTTTGCCAA CAGCCAATAT CAGCACGTTA CATGCCTTTT GTTAACTGT TATTCGTCGT	8760
TTTTATTATT TAATTGACAT AGATCCAGTT TTTCGGATGT TAACAGATGA AACAGAAACA	8820
TTGTTATTAA AAGAAGATGT TTGGGATGCA TTACGTGAAC AATTTTATGC GGAAAATCAA	8880
GAAGCGTTTT ACCAATTAAC AGCGAATTTT TCCAATGACC GTAGTGATGA CGGACTCACT	8940
AACTTAATTT TTTCTTTTGA TGAATTTGCC AAAGCCAATC CTGATCCAGA GGCGTGGATT	9000
AATGGCTTAA CACAAGCCTA CGAAGTAGGG GaTCAGCTAG GCGAATCAAA GTTATTTTCA	9060
ACGTATTTAA AACCGCTAGC CGTTGAGACG TTACAACGAA CTCTTCAGCG TTACGAAGAA	9120
ATGGTGACCT TAACAGAGGG CGAAGAGAAA CTTCAAAGA TTTGGTATTT AGCTCmAAAT	9180
GAAAAGGAAC AAACAAAACA ATTTTACAA TTTTATAGA GAAATGATTT AGAGAGCGCC	9240
TATAATTTGA CTGAATTGTT GAGCTTTGAC CGCTACCCAA CGGTTCTGTC CGAAGAGTTG	9300
AAACCAACCG CGGAACAAGC CAAACAGTTA CGAGAACAAA ATAAAAAAGC GTTAAATGAT	9360
TTGAAAAAGC AACTGTTTAC ACTTTCCCA GATGCCATGA AACAGTCTT GAAAGAGGCC	9420
ACGCCCATTG TTCAAGAAAT GGCTCACGTG GGGAAACAAT TTATGGAGGC CTATGGCGCA	9480
GAAAAACGGC TTAAAAATTT AGTGGATTTT AACGATTTAG AACATTATAC GTTAGCTATT	9540
TTAGCGAAAA ATCAGGCAGA TGGCTGGCAT GCATCAGAAG CCTCGGTTTA CTATCGCGAA	9600
AAATTTGATG AAGTGTTAGT CGATGAATAC CAGGATATCA ATCAATTACA AGAATCTATT	9660
TTGTATTGGT TGCGCCGTCC GCTTAGCACA GAAGGAAACC TTTTATGGT AGGAGATGTG	9720
AAACAGTCTA TTTATTCTTT CCGCTTGGCA GATCCTACAT TGTTTATTGA AAAATACAAT	9780
CAGTATGGGC AAGGCAAAGA GGGCAAACGA ATTATTCTAG CTGAAAATTT TCGTTCTCGC	9840
AAAGACGTTT TAGACTTTAC GAATTTAGTA TTTAGTCAAT TAATGGATGA GCGAGTGGGC	9900
CAAATGCTT ACGATGAGTC AGCTGCTTTG GTTCATGGTT TTGATCAATT TTCGGAGGCG	9960
GCTGATTATT CAACTGAATT GCTAATTTAT GAAAAAAAG CAACAGAATC AGTTGAATTT	10020
CCAGAATTAC AATCCCCTGA GTTGTGATT GAAGATAAGA CAGAAGGGGA ACTTTACGTT	10080
ACTGCTTTGA AAATTCGTGA ATTAATTGAC CAAAATTTCT TAATTTATGA TAAAAAGCTG	10140
AAAACGGATC GACCGATTAC TTATCAAGAT ATTGTCTTAT TAACGCCTAC GAAGAAAAAT	10200
AATTTAACTA TTTTAGATGT TTTTAAATCA TTAGAAATTC CAGTTCAAGT CAATGATGCT	10260
CAAAATTATT TCCAAGCAAC GGAAATTCGT ACAATGATTG CCTTGCTTCA ATTGATTGAT	10320
AATCCTTATC AAGACATTCC TTTGGCAGCT GTGTTACGTT CACCAATTGT TGGATTAAAA	10380
GAAAACGAAT TAGTGCTTAT TCGTTTGGCT AATAAGGAAA CCAGTTACTA CGAAGCATTT	10440
TTAACGTTTA ATCAAAAGAT GGAACCAACC ATGGAGGAAG CAGTTGTTCA GGAAAAAACT	10500

ATACGTTTTG CTGAATCATT AGAAAAATGG CGGGAACAGG CTGGACGCAA CCAGATTAGC	10560
AACTTGCTGT GGACGATTTA TCGGGAAACG GCCTATCTTG ATTATGTTGG CGGGTGCCT	10620
GTTGGCAAAC AACGCCAAGC CAATTTATAT GCGTTGGTCG ATCGGGCCGC GCGGTATGAA	10680
AAAACAACCT TCCGCGGATT ATTTCAATTT GTTCGCTTCA TCGAAAAAAT GCAAGAAAAA	10740
GACAAAGATT TAGCAGAAC AGTCGTTTTA AGTGAAGAGA ATGCCGTACG CGTCATGACG	10800
ATTACGCTA GTAAAGGGTT AGAATTTCCG GTGGTATTCG TTTTAGATAT GACCAAAGAA	10860
TTTAATGTCA GTGATTTGAA TGAACGCTAT ATCTTTGAAG AAAATTTAGG GGTAGGCATT	10920
CGTTATTTAC AACCTGAAGA ACGAGTTATG TACGATACTT TACCATTTTT GCGATTAAA	10980
CAAGTACGAT TAAGAAAATT ACTATCCGAA GAAATGCGGA AGTTATATGT GCGGTTGACA	11040
CGTGCAGAAC AGAACTTTTT TCTGGTAGGC TCCTATAAAG ACCAAGCGGC GATGTGGAAA	11100
GAGTGGCTGA AAGTTGGCGA TGTAAGAAAC CTGGTGCTAC CAGCTGAAAA TCGCTTGCAA	11160
AGTAAAAGTA GTTTAATGAA TTGGGTTGGG ATGACATTGG TACGTCATCA AAAAGCCGAT	11220
GAATATCAGC AAGAAGTAGT AGTAAGTAAC GTTCCGCAAG TCAAAAAGCA CCCAGCTAAT	11280
TTTCATATTC AGTGGTTCAA CGAAGAACAA CTACGAGCAG CCATTTCAGCA ATTACAATTG	11340
CCAGAAAGAC AAGCAGAGGA TTTAGCCGAA AAAGCGCAAT TATCTGCTGA CAAAATCAAT	11400
CGTGGCCTTG CACGGTTAAG TTTCAATTAT CCGTTTGAAG TGGCTACTCG GACCACCAGT	11460
TATCAGTCTG TTTCTGAAAT CAAGCGAGTA TTTGATGACC CCGATAACAA AGAAATTGGT	11520
AAAATTGAGG TTCGGGAAGA CAATACTATT CAAGCGCAAC CGTTGATTGT CAATCGAATG	11580
ATTGAAGGAG ATTTAAGTAA GCCTAAATTT TTAGATACGG TTCAAGCGCC CTCAGCAGCT	11640
GAAATTGGGA CAGCAACCCA TTATTTATTA CAACTAATTG ATTTAAGTAA GCAACCTTCT	11700
TACGAGGAAG TGCGAGCTGT TCAAGAACGC TTAGTGGAAG ATAAATTGAT TTTACCAGCA	11760
ATTGCTGAAA AAATGAATCT GGAACAAATC GTGGCCTTCT TTGACACCGC CTTAGGAAAA	11820
CAATTGATTC AACACCATCA AACC GTTAGA AGGGAACAAC CTTTTTCTAT GCTGATTGAA	11880
GCAGAAGAGC TCATTCAAAA TTATCCAGAA ACGACACAAG ATGATTTATT AATTCATGGG	11940
ATTATTGATG GCTATATCGA ATTAGACAAT CAATGCATTT TGTATGACTA TAAAACCGAT	12000
CATGTAAAAG GCACCTCTCC CCAAGCAATT TCTGAAATTG TCGAACGGTA TCGCGGACAG	12060
ATGAATCTTT ATCGTAGAGC GCTACAAGAG GCAACGCATA AAGAGGTGTC GCACGTCTAT	12120
TTAATTTTAT TAAATGGAGG AGTAATTATT GACATGCAGA CmGGAAATGT TGTAGATTTT	12180
ATAAAGTAGA AAAAATTAGA AGCAGAAACA CTTACATTTT GTAAGGAATG TGCTATACTC	12240
ATTGAAGAGG TGAGGAAGAT GGAAAAAGTA AAAACAGAAT CGTTTGTACT TTGTsCCgaA	12300
ATTTGAAAAA TCATTTTCAA TACTTGGTAA AAAATGGAAT GGTTTAATTA TTGATGTTTT	12360
ATTAGAACGA GGTCTCTAAC GTTTTGGCGA ATTAAAAGAA xAAATTCCAA TGTTAAGTGA	12420
TCGTGTCTTA GTAGAGCGAT TGAAGGAAC AGAAGCTGAA GGAATTATCA CCAAAGCTGT	12480

TCGTTGCGGT GAAGGAAATC GCTTTGGAnT ATTCCTGAC CGAAAAAGGC GAAGACTTGC	12540
AATTAGCAAT GGAACAAATT CAACATTGGG CAGAAAAATG GGTAAAAGAC GAAGAATGCA	12600
GTTGACCGCA AGCAAAACTT CTTGTAAACT AAACGTAATT AAATAACTAA AAAAACAGTG	12660
ATGGAAACAA GTAGTTTGAA TTGACGTGTT TTAGGGAGAG TCAGCCAGAG ACTGGAAGCT	12720
GATTACACA GTCACCAAGc GAAATTCACT TCCTGAGTTT GTTTGTTTTT AAGAAGAAAG	12780
CAAACCGTA ATTGCCGTTA AATGAAATGA GTGCATGACA ACAATTTGTT ATGAATTAGG	12840
ATGGTAACAC GATAGCTCGT TCCTTTGAAA GGAATGGGTT ATCGTTTTTT TTTTAGTTTA	12900
AATAAAAGAT AGAGGGGAAC GCATAATGAC ATTACAAGCG CAATTAGAAG CTCTTAGAGA	12960
CAATACGCTC AAAGAAATCG CACAAGTTGC TACTTTAAAA GAATTAAACC AAATTCGCGT	13020
AGAAACATTA GGGAAAAAAG GGCCAATCAC CGAAGTTTAA GAGGCATGAA AAACCTTTCA	13080
CCAGAAGAAC GACCGGTGGT GGGGGGCTTT GCAAATGAAA TTCGTGATTT ATTAACAGAA	13140
GCAATTGAAG CGCGCAAAGT TGTTTTAGAA AACGAAGCGT TAAATGCGGC ACTAAAAGAA	13200
GAAAGTTTAG ATGTAACTTT ACCTGGGAAA CAAATGCCTC AAGGCACACG TCATATTTTG	13260
ACACAAGTTA TGGAAGAGAT TGAAGATATT TTCTTAGGCA TGGGTTATCA AGTAGTGAA	13320
GGATACGAAt AGAaTCCGAT CATTATAATT TCGAACGCAT GAATCTACCA AAAGATCATC	13380
CGGCACGCGA CATGCAAGAT ACGTTTTATA TTTCAGATGA GATGTTAATT CGTACGCATA	13440
CTTCACCACT TCAAGCTCGA ACAaTGGaAA AACACGATTT TTCmAAAGGk GCATTACGAA	13500
TGATTTCTCC TGGGAAAGTA TTCCGTCGTG ATACAGATGA TGCGACACAT AGTCACCAAT	13560
TCCATCAGAT TGAAGGCCTT GTTGTGATA AAAATGTCAC AATGGGTGAT TTAAAAGGGA	13620
CGTTAGAAGT CATGATGaAA AAAATGTTTG GtGaAGATCG TAAAATTCGT TTACGTCCTA	13680
GCTATTTCCC yTTTACAGAA CCTTCAGTCG AAGTAGATGT TAGCTGTTTT AAATGTGGCG	13740
GCGCAGGTTG TAACGTCTGC AAACATACAG GTTGATTGA AATTTTAGGT GCGGGGATGG	13800
TTCATCCAGA TGTGTTACAA ATGTCAGGGA TTGATCCAAC AGAGTACTCA GGCTTTGCCT	13860
TTGGCTTAGG ACCAGATCGC GTTGCTATGT TACGTTACGG TGTAATGAT ATCCGTAATT	13920
TTTATCAAAA TGATTACGT TTCTTAAATC AATTCAGGT AAAGGAGTAG TTGAAAAATG	13980
TTAGTTTCTT ATAAATGGTT AAATGAATAT GTTAATCTTT CAAACGTTAC TCCGCAAGAA	14040
TTAGCGGACA AAATGTCTGT GACAGGTATC GAAGTTGAGG GCGTTGCTGT ACCTGAAGAA	14100
GGATTAAAGA AAATGTTGT CGGCGAAGTA AAAGAATGTG TTCCTCATCC AAACCTGAC	14160
CACTTATCAA TTTGCCAAGT AGACATCGGT GAAGAAGAAT TGTCACAAAT TGTTTGTTGA	14220
GCACCAAATG TAAAAGCCGG AATTAAAGTC ATCGTAGCAT TACCTGGTTC AAGAATTGCA	14280
GGAAATCAAA AAATTAATAA AGGTAAATG CGTGGCGAAG TCTCTAACGG AATGATTTGT	14340
TCGTTAGAAG AGCTAGGATA TTCAGATAAT GTCGTACCAA AAGCCTATGC TGAAGGGATT	14400
TATTATTTAC CTCAAGAAGC AGTGAATGGG ACACCTGTTT TCCCTTATTT AGACATGGAT	14460

GATGCGATTA TTGAATTATC AATTACACCA AACCGTGCAG ATGCACTAAG TATGAGAGGG	14520
GTAGCCTATG AAGTTGGCGC AATTTATCGT CAAACACCTC AGTTTAATGA TCCCGAACTC	14580
AAAGAAGATG CTTAGATAA CGTGGAAAAT TACGTAACAG TGACTGTTGA AGATTCACAA	14640
GATGCGCCAg cGTATCAAAT TCGTGTTATT AAAGATGTGA CGATTGCAGA AAGTCCTCAG	14700
TGGTTGCAAA ACCGATTGAT GAATGAAGGA ATCCGTCCGA TTAACAATGT GGTGGACGTG	14760
ACAAATTATA TTTTATTATT ATTTGGTCAA CCATTGCATG CATTTGATTA CCAAAAATTA	14820
GATAGCAAAG AAATCTTGGT TCGTCGAGCA ACAGCAACAG AAGAACTAAT TACATTAGAT	14880
GGTGAAACAC GTCAATTAAC AGAAGAAAAT ATCGTCATTA CGAATGGAAA AACgCCTGTC	14940
GGCTTAGCCG GAGTAATGGG TGGGGCTAAT TCTGAAATCA GTCAAGAAAC AACAACTGTG	15000
GCTTTAGAAG CGGCATTGTT CAATCCATTG TCTATTCGCA AAACGTCTAA ACAATTCAAT	15060
TTACGGAGTG AATCTTCAAG TAGATTTGAA AAAGGCATCA ACCAAGCGAC AGTTGGGCTA	15120
GCTTGTGATG TGGCAGCCGC GATGATTGCA GAGTTAGCCA ACGGAACAGT GGTCTCAGGC	15180
ACTGCAGTTG GTTCTGAAGT GGCCGTAAAG GAAGCGCAGG TAGCTGTGAC TTTAGAGCGA	15240
ATCAACCAAT ATTTAGGCAC TGCATTAGAT GAAGCAACGG TGAATGAAAT TTTTGAAGCA	15300
CTCGGTTTTG CTTATGAAGT AAATCAAGGC GCATATGAAA TTACGATTCC ACCAAGACGT	15360
TGGGATATTG CGATTGAAGC AGATATTATT GAAGAAGTGG CGCGCATTTA TGGATATGAT	15420
CATTTACCTT CAACATTGCC AAGTGGAGAA ACAGTTGCTG GAAGTCTAAC CAAAGCACAA	15480
CATGTTACGC GCCAATTAAA GAGCTTACTG GAAGGTCATG GTACGAGTGA AGCCATCAGC	15540
TATGCGTTGA CAACAGAAGA AAAATCCCGT CAATTTATGA TGAAAGAAAG TCAAACAACA	15600
CGTTTGCAAT GGCCAATGAG TGAAGAGCGT TCTGTGTTAC GTATGAACTT AATTTCTGGT	15660
TTATTAGATG ATGTCGCATA CAATGTGGCA CGTAAAAATA ATAATATCGC CTTCTACGAA	15720
GTAGGACGCG TTTTCTACCA AACAGAAGAT CCAACAAAAA ATTTACCTAC AGAAGAAAAT	15780
CACTTAGCAC TTGCTTAAAC TGGAATACA ATGGTTAAAG ATTGGCAAAC AAAAGCAACA	15840
GCCGTTGATT TTTATACAGT AAAAGGCTTA GTGGAAAGTA TTGTAGCTGT TTTAGGATTA	15900
ACAGAAAAAA TCAGTTATCA AGCGACAACG GCGATTCCAG AAATGCATCC AGGCCGGACA	15960
GCATGGATTT ATTTAGAAGG TGAAGTGGTT GGTTTTGTG GACAAGTTCA CCCAACGACA	16020
GCGAAAGCGT ACGATATTCC AGAAACATAT GTTGCTGAAT TAACTTGCa ACAATTAGTA	16080
GCCATAGAAG CAGGCGGGGT TACTTATGAA GCAGTTTCTA AATCCCAGC AGTTTCTCGA	16140
GACATTGCTT TATTAGTCGA TGAAACAGTT ACCAATCAAG AACTAGTTAA AACTATTTCA	16200
GATAACGCAG GTAAATATTT GAAAGAGATT CATTTATTTG ATGTATATCA AGGTGAAAAA	16260
CTAGGCGCTG GCAAAAAATC AATGGCGTAT AGTTTAACTT TCGTTAATGC AGAAGCGACA	16320
TTAGTGGATG AAGAAATTAA CCGTTCAATG GAAAAAGTTG AAAAAGCGCT AATTGAAAAA	16380
CATCAAGTAG AAGTAAGATA AAATAAAAAC AGGAAAGTGG GGCATAAGTC GAAATGACTT	16440

ATGCCCCACT TTCTAAATCT TAAGATAATC AGCTGAACAA AAGCAGCTCC AAGAAATAAG	16500
CCGAAATTCT CCAAAAATTA AAGAACAATT TTCGGAAATT CCTTCTTATT TCTTGGAGCT	16560
AAACACTTTT GTCTCAGCCT CCTATTTTAA ACTACATTAA TAACTTTTCG CCCAATTTTCG	16620
TGAAAGCCAA CGTGAAAATT GTGAGATACT AAAGTTAATA ATGAAATAAA TTAGGGCGAC	16680
GATTCCGTAA ATAGTGAATA CTTGTCTGT TGTGCGTA CGCCCCATTA AGATATAAGA	16740
TTTGCCAAAG AGTTCTTAA AAGCAATCAC GGAATACAAG AAAGTTGTAT CCTTGATGAC	16800
AGTAACAAAC TGAGAAACAA TTGCTGGTAA GACATTGCGA ATGGCTTGAG GCAAAACAAT	16860
ATGATAAAGA ATTTGCCAAT TTGAAAAACC TTGAGAACGT GCCGCTTCAA TTTGGCCTTT	16920
GTCGACCCCA TTAAACCAC CACGAATAAT TTCCGCTAAA GCAGCGGTGG TAAACACAGT	16980
AAAACTCACA ATTCCAGCAG GTGTGGATTT GATTTTGAAA ATCAAAAAAA TAACATAAAT	17040
CCACAATAAG TTTGGAATAT TACGCACCAC TTCTACATAA ATACTGGCAA GTAACCTCAG	17100
TGGTCCATGG GTTTGATTTC GCAAAACGGC TAGAATGGTG CCAAAATCG TACTCAAAAT	17160
AATTGAGACG AAGGAAATAT AGAGCGTCAA TTTTAAGCCG TTAAATAGAA ATTGAAAATT	17220
ATTTGCTGTC AGTAATTGAC GCATTCTTC TGAAAAAGGC ATCTGTTTTC CTCCTATCTG	17280
CTATATGCTT TCTTATTTTT TTCTTCATT TTTCTAGCCA AATTCGCTAA TGGGAAGCAA	17340
AGAATGAAAT ATAAGAAACC AGCAAGCGCA AAGGCGGGAA TATAGTTCAA GTTGACAGAC	17400
GACCAACTAT TGGCAGCAAA CATTACGTCA GCGCCAGAAA TAATGGCAAC GTTGACGTA	17460
TTTTTAATTA GGTCACTAC TTGATTGGTT AAAGGCGGTA ACATAATCCG CCATGCTTGT	17520
GGTAAATCA CATACGCAT CGTTTGACCG TACGTTAATC CTTGGGAGTA AGCCGCTTCA	17580
AATTGCCCTC TTGGGACAGC TCCGATTCTT GAGCGAACCA CTTAGAAAT GTAGGCGCCA	17640
TGATAGAGGC CTACACACAG AATGGCAATC GTAGTAATAG AAAATGAAAG CAGCGGGCTT	17700
AGCAGAGGAA AGCCATAATA GACTACGATA AATTGAATCA ACAAAGGTGT GTTTTGAAAA	17760
AATTCTACAT AGACACGACT AATTGCGTGT AAAAGTTTAC TTTTGGTCGC TGATAAACTG	17820
CCAAAGAAAA TGCTTAAAGC AAAGGCAACT AAGATCGCTC CTAAAGAAAG CAGAATAGTA	17880
TAGAGAAATG CCTGACCAAA AATTCGCCAA TCTTGAAATA AAGCCTCCCA ACGATATAAG	17940
GCAAATGGCC CTTCTTCAAG CCAGGAAATA TTGGCTAACA AAAACATAAA AATCCCTCCT	18000
AAAATTTTGT ATTCTTCAAG TCTTAGGAAT GATTAATTTT TTAAGCCCCA GTCTTGATAA	18060
ATTTTTTCTA GTGTGCCGTC TTTTCCCAT TTTTCAATTA AACCATTTAG ATAATCATTT	18120
AATTCAGTAT TTGCTTTCTT TGTGGTAATT CCATATTCTT GTGGCGAAAA ACCATCTTTT	18180
AAAATTTCTG TCTTGCCATC CACATAGCCA GATAGGATTG ATTTATCCAC TGAAATGCA	18240
TCAATTCGTT TAGAAGTTAA CGCTGTTTTT AATTCTGGAT AGGAACCTAA TTCTTGATAA	18300
GTAAATTTGA TTCCTTCCTC TTGCGCTTTT TTCTCTAAGT TTTCTTTTGT CGTAGCAGAT	18360
TGTGCAACGC CGATTGTTTT TCCGTCTAAA TCTTTCATGC TGGAGAAGTT ATCCTCTTTG	18420

CGCACAAAGGA AACCAATCTC ATCTGTATAA TAAGGAGTTG TAAATTATA AGTCTTTTAA	18480
CGTTCATCAG TAATTGTAAA TGTAGCAATT ACCATATCTA GTTCACCATT GTCTAAAAGG	18540
GGACCGCGCG TTTTCGCTGT CACGCCCACA AATTCCACAT TCTCTCGCT GCCGTAATT	18600
TCTTTCGCAA TCAAGCGTGC AATATCTgGT TCCATTCTT CA <del>t</del> TTTTGTn TGATCAGGAT	18660
TCATATAGCC AAAATTAGGA ATGCTTCTTT GACACCTACT TnAATnTnCC CTGCTTTTTT	18720
TAAT	18724

## (2) INFORMATION FOR SEQ ID NO: 58:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7254 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

AAAAATTACG TGATTCGCTT AGCCTGACCA GACTTTTGTC TAAGTCTTAA AACTTGATAA	60
GTAATAAGAG GTCGGGACAG AAGTGCTTAA CTCCGAGAAC AAAGTAGGTA CTGTTCAACA	120
TCTATTTGCT TTGCAAATAG TGTTTCTCAG CAATAAGAAG AAATTTCCGA AAATTGTCAT	180
AAAACACAAC GAGGAACGAG TTGATGTTGC ATGATACACT ACACTCCGTT TCGTTCGCAT	240
CAATTTCTAA GCATGGAACA TGCTAGAATT GAAGGACCCG GTGTTCTTTA ATTTTGGAG	300
ACCAAGTCCT TCAATTCTTA GTGATTTATC ACTTAGAGAT TGATGAGATC GGAGCTCTGC	360
GAAGTAGGTA TTGTTTCATCA TCGAATCACT ACGTTCTTCG TGATTCTCAA CAATTTCCGC	420
TTATTTCCGA AGGACCTTCA CCAAGTAGGT ACGATGATT CATCTGTTCA TGCTTCACAG	480
TGAAACATCG CAACTCTTAG AGCTGTAGCT CTTTGAAG CTGTACTGTG ACGAGACGCA	540
TGTCGAGAAA ACACCTTAGA AGTTGTTGAG ATCGGAGCAA CGCGTAGTGT TGCTTCTGTC	600
CCCCCGTTT ATCAATTTT AAGCCTGGAA CAAAAATCCA AAGTGATTTT TGTTCCAGGC	660
tCTTACTTTC ATACTAATCT TAGGTCAGGT TGTCATCAAG TTTGCCAATA ATAGACTTTT	720
ACAAGCGCGC CTTTTTTGAG TAGTACACCA GCTGGTTGTT CTAGCAGGCA ATTTGTTTCT	780
AAAGTATTTG ATAAATTGGA TGAAAGATGA CTTTTGTCA CTAAATTTAC TGTGCCATCT	840
ATTTCTTTCG CTCGGTAAAA TCGACGTAGT CTAAGGGG TTAATTCATG CGCTAAAATT	900
ACTGTTTCGTT CCTTGAGATT TAACTCGACA CAATTAATAA AAGTAGCTAA TAATGTCCAG	960
TAAAAAAGAT GTAAATTAAC CATTGCCGCA AACGGATTGC CTGAAACGCT TAAAATGAGT	1020
GTTTCTTGAA ATTGACTTCC CATAACTGGG GTTCCAGGTT TCATATTTAC AAAATGAAAC	1080
AACTCTGTTC CTCCGATACG TTTAATTGCT AAAGGTAAAT TGTCCTTTTC ACCTACCGAA	1140
ACACCACCAC AAGTAATTAA TACATCAAGT TTTGGTGAAA TCTCAATCAG CCTCTTCGCT	1200
AATTCTTCAG GGTCATCTGA ACAATGGTCT AGGTGGATTA CTTCAATACC ACTCGCTTGT	1260
ACAAAAGCTG CTAACATGTA GAGATTACTA TTATAAATTT TCCCAACTGA CAGTGGTTCT	1320



TCCAAAGAAA CTAACCTCACT ACCGGTCGCT AAAATCCCCA CCTTCATGCG CTTCAAAACA	1380
ACAACCTTAT CAATCCCTTG CATTGCCAAC AAACCAATCA TGCGACTATT AATCTGTTGA	1440
TGTTTTGCTA AAATTTTTTG CTGAAAAGCA ATATCTTCTC CAATTTCCGC ATAGTTTTGG	1500
CCCGCAGTGA CAGAGCGAAA AATTTTAACA TGGTCTTGCC CAAAATCCGT CCATTCTTGT	1560
GGCACTACAG CATCAAAATC GGTGGAATA AAAGCACCTG TTGTTATTTT AATCGCACAG	1620
TTATTCGTCT GATTACAATA AAACATGGAA TCCCCTGCCA CTAAAGAACC GACGACTTTT	1680
AAACAACCTG GGCTTTCAGC AGTCGCACCT CGTGTCTCTA CTGCACGAAC GGCATACCCA	1740
TCCATCCCGG CACGAGGAAA ATGAGGAATC GCTGTCTCTG CAAAAATGGT TTCTCCGCTA	1800
ATCCGTCCCA AGGAATCAAG CAGCGCCACT TCTTCTGTTT CACATTGACT CTTCATATGT	1860
TGGGTCATTT TCTCTCGTGC TTCTTCCAAC TCGATCATGT TTTAACCATC CTTTTTTATT	1920
TATTCACCTG TTTAATCGCA ACGGCACAAA CTTTATATTC AGGAATTGTG GCGATATCAT	1980
CAAATACAGC ATTGGTGATT TCGTTAACAT TCCCATCTGG AAAGTGAAT GTCATAAAGA	2040
CTTCTTGCGG AAATACTCGA TTTCCGACAG CCGCATAGGT TTCGATTTTG CCACGTCGAG	2100
AATGAACCTC AACTTTGTCA CCTTCTTGGA TTCCCAGAGC TTGTGCATCG ACTGCATTTA	2160
TCTCAATATA AGAGTGATTC GCTATTTGGT TAATCCCTTC CGTTCGTCCT GTCATTGCTC	2220
GGGTATTGTA ATGGTAAAGC ATTCGGCCTG TCGACATTAA ATACGGGTAC TCTTCATCTG	2280
GTAATTCTTG TGCTTGCTTA TACGGAATTG CTTTAAATAA ACCTTTCCCA CGAGCAAAAC	2340
TACCCACATG CATAATCGGC GTACCTGGAT CTGTCAAGCT CCGACAAGGC CATTGTAAGC	2400
TCTCCTTCTC TAATCGTTCA TAATTAATTC CACCAAACGA AGGCGTCaCA GCTGAAATTT	2460
CTTCCATAAT TTCTTTTGCA GATTACATCG CACAAGGATA GCCCATTCGT GTCATGACTT	2520
CACAAAAAAT TTCATAGTCT TCTCGCGCTT GTCCTCTTGG TTCAACAGCT TTACGGACTC	2580
TTTGCACGCG CCGTTCTGTA TTTGTAAATG TCCCCTCTTT TTCTGCATAA CTGATGCCTG	2640
GTAAAACGAC ATCTGcATAA GCAGCCGTTT CGGTCATAAA TAATTCTTGA ACAACTAAAA	2700
AGTCnTAAAC TTTCTAAAGC TTGACGAACA TGTCTGTGT CTGGATCCGT AACAAATTGA	2760
TCTTCTCCAA AAATATAAAG GCCTTTGACG CTTCTGCGG TTGCTGCTGG TAAGACTTTA	2820
GTAGAAGTTA AACCTGTGTT TCGATTTAAC GGCACATGCC AAGCTTTTTT AAATTTGTCG	2880
ATCACTTCTG GATTATTGAC TTTTGTATAG CCTGGGAAAT CATAAGGCAT ACAGCCCAT	2940
TCGCAAGCAC CTGGAACATT ATTTTGACCT CGCAAAGGAT TGACACCACA GCCTGGTTTT	3000
CCAACCTTAC CAACCAACAT TGCCAAATTA GACATACTCA TAACACCCTC AGTTCCTGTC	3060
GAATGTTCCG TTACACCTAA ACAATAAATG ATTGGTGCTT TCTCTGCTTT TGCATACATG	3120
CGTGCTGCTT GAATTAAATC TTCTGGATGA ATATGACAAA TTTCTGCGAC TTTTCCGGC	3180
GTATAATCAG CCACCATTTT TTCTAAATCT AAAAATCCTT CTGTCCGTTT TTCAATAAAA	3240
TGGCGATCTG CTAAACCTTC TTTCAAAATC ACATGCATCA TGCCATTAGC AAATGCGACA	3300

TTCGTTCCCG	CCTGAACTTG	TAAATGCAGC	GCACTATCCT	TCACTAAGTT	AATTTTCCGC	3360
GGATCAACAA	CAACAACCTG	CGTTCCTCTT	TGGATGGCTT	GGCGAATTG	CGCACCAATA	3420
ACTGGATGAG	CTTCTTCTGG	ATTGGAACCT	ACTAATAAAA	TCATGTCGAC	ATCTTCAGTG	3480
ATATCAGCAA	TAGGATTAGT	CATCGCACCA	GAGCCTAAGG	TTTGGGCTAA	ACCATGGACC	3540
GAAGCCGAAT	GACAAACACG	GGCACAGTTA	TCAACATTAT	TGGTACCAAA	AGCAGCACGC	3600
ACCATTTTTT	GAAAGACATA	ATTATCTTCA	TTTGTGCTC	TAGAGCACGA	AAAACCAGCA	3660
AGAGCATCCG	GACCGTTTC	TGCTTTAATT	TCGTTAAATT	TTGACGAAAC	AAGCGTTAAG	3720
GCTcTTCCCA	AGAAGCAGGT	TCAAAAATGC	CATTCCGTTa	ATTAATGGCT	CCGTTAAGCG	3780
ATCTCCTGAC	CCAACAAATT	TATATGAAGC	AAATTTTCCT	TTCACACAAA	GTAAATTTTT	3840
ATTAGCTGGT	CCATCCAGGG	GTTCAACACC	CACCAAACGA	TTATTTTTTAA	CGAGTAAATT	3900
CATCTGACAG	CCTGTCCAC	AATGCGGACA	AGTTGTAGGA	ATTTTCTGGG	TTTCCATTT	3960
GCGGTACTCT	kTCGTATCTT	TTGCTGTTAA	CGCACCTGTA	GGACAGGAAG	AGACACAATT	4020
GCCACATGAT	TCACAGATAG	ATTGATCAAA	CGCCTGCCCA	TAACTCGGCA	TCATTTTTGT	4080
TTCAAAGCCA	CGATTAGCAA	TACTAAGTAC	ATCTCTTCCC	TGCCGCAATT	GACAGACACG	4140
GGCACAAACGA	CGACACATGA	TACACTTTTC	TGGATCATAA	GAGAAAAATG	GATTACTGGT	4200
ATCTTCTTGA	TGGCAAGGCA	TTCTTTTTTC	CTCGGTAAAG	CTAGTGACAT	CAATCCCATA	4260
ATCTAGCGCA	TACTGTTGTA	ATTGACAATC	CCCGTTTTTT	CCACAAGAAA	AACATTCTAA	4320
TTTATGATTA	CTCAATAATA	AATCAAGAAT	AAAACGTCTG	GAATCAGCAA	CTTTTTCAGA	4380
ATGTGTTGCC	ACAACCATGT	CCTCTTGACA	ATGTGCCGTA	CAAGCAGTTG	TTAAGCCACC	4440
ACGTCGGCCA	CCTTCCACCT	CAACCACGCA	CATCCGACAA	GAACCATCAG	GTGCTAATTC	4500
CTTTAAATGA	CAGAGCGTTG	GGATTTCAAC	ACCTAGCCCT	TTCGCTGCTT	CGAGAATGGT	4560
AGTCCCTTTA	GGCACAGTTA	CTTCTTGGTT	ATCAATAGAC	AACGTAAC TG	TTGCGTATG	4620
CAACTTTGTT	TTCATTTTTA	CCAACCTCCT	TTGCTGGTGG	CATAGATTTT	TGTTTCAGCA	4680
GTCGTTTCAA	ATTCTTCTGG	AAACAAAGCC	AATGCACTCT	TCATTGGATT	tGCGACGGAT	4740
TGGCCCAAAC	CACAAGCTGA	TAATCGCGTG	ACATGTgTGA	GCATCTTTTc	cAaTCGCGGC	4800
AAATCACTAG	CTGGTTGCCT	CGTTCCGATk	GAATTTACTC	AATAATTCCA	AAATCCTAGT	4860
TGTTCCCAAT	CGACAAGGGG	TGCATTTACC	ACAGGACTCA	TGAGCAAAAA	AGGCCGCCAC	4920
GTGTACTAAA	TAGTCCACAA	CATTAAGTGA	ATCATCCATT	ACTACAATTG	CACCAGAGCC	4980
GACAGATAAG	TCATGGGCCC	ATAAATCTTC	GTAGGAATAC	AAACAATCCC	TTAACGTTGC	5040
GACAGCACCA	ATTGGACCGG	ATTGTCCACC	GAAGTGAATA	AATTTCAATG	CTCGACCAGT	5100
TGCAGAGCCT	CCGCCGTATT	GCTCGCCATA	AATTATTTCT	TCTAATGGTG	TTCCAAGGTT	5160
GACCTCGTAT	AAGCCCCGAT	TTTTAATATG	TCCAGACAAA	CAAATTAGCT	TTGTTCCGCC	5220
ACCATCGGCA	GTACCCATCT	CTAAAAAGGC	TTGTCCACCT	TCTCTTAAAA	TTACAGGAAC	5280

ACTTGCAAAT GATTCAACAT TATTTACTAA TGTTGGTTGC AAATATAAAC CAACATCTGC	5340
TAAGTGTGGT GGTTCGACAC GCGGCCGCCC CGTTTTTCCT TCAATTGAAT TTAAAAGTGC	5400
TGAATTTTCC CCACAAATAT AGGCGCCAGc ACCAGAAATA ATCGTAATAT CATAATTGAA	5460
TCCTTCAATT CCTAAAATGT TTTCCCTTAA AAAGCCAGCT TGTCGTGCAT TATCGAGTGC	5520
TTCTTGAAAT GTTTTTTGGA TTCTGCGATA TTCTCCTCGC ATATAAATAT AACCTGCTTT	5580
TGCTGAAAAT AAATAGCCAG CAATAATCAT TCCTTCAATC ACACTCAATG GATCTTCAGA	5640
TAAAAGTACT TTATCCTTAA AGGTCCCTGG TTCGCCTTCA TCAGCATTAC AAACAATGTA	5700
CTTTGTGGtT CCTTTTGcAT GATAAAGATG GcGCCATTTT TTTCTTAAAG GaTAGGcTGc	5760
TCCACCGsGa CCTCTTAAAGT GcGcAATATC CAGTTCGTTT AAAATTTCTT CATCCGATAG	5820
TGAAATTGCT CTTTTTAAAC CATCAAATCC ATCGTATTTA CAATATTCAG CAACATCgGT	5880
CGCAGATTGC ATTTTCCCCA TTCTCTCTAA TAACATTGGT TGGTTTCTCT TAAGCATTGA	5940
CTCACCTCyT ACAGTCCTTC ATAGCAATCA TTTTGTAAT GTTGAAWTAA TTGATAAATT	6000
TTCTCTTCTG TTAGTTGACT AAAAACGGTG TCCTTAATTT TAATTACTGG ACCTAAATCA	6060
CAAGCCCCGA TACAAGGAAT ACTATGATAC ATAAATAAGC CATCTGGTGT CGGCTGATTC	6120
TCTGGCACTT CTAAAATGGT TTCTAAAAC TCGGCAACCA TCGCCCCACC AGTATAATGA	6180
CACGGTGTGC TGTTACAAAT CTTCAAGACA TACTTCGCTT GCGGTTCCGT CTTTAAAATG	6240
GCATAAAAAC TGACGATTTT ATAGACTCGC GCTTCCGTTA AATGTAGATG CTCTGCAACT	6300
AATTGAGCTG TCTCTGATC GATATACCCT TCCTCTGAAG CAAATTGTAA TTCAATTAAA	6360
ATATTTAAAA TTCGTTGTGG ATCAGCATCG TTTTCTAGAA TAATCGCTTC TTTTCAATT	6420
AAGCTTAATG CACACATTCC TTTTATTCC CCCTTTTtag TGAAACTTGA AGTTGTTCTC	6480
CTTGATAATA GATACGCAAT TTTTTTGGTA TTTCAGCAGA TAGTAAACAA AAATCGGCTT	6540
TATCTATGAC TTTTTTTGCT GATTCTTTTT GCGGATTCTG GTTATTCATC ATAAGAAATA	6600
GTCTGGCTG CACAACATTA CGTAATGAAT TTGATTCAAA AATGATTAAT GCTTGGGCCG	6660
GCACAGCTTT CAGAAATTGC TGAATGGCTT CCGCCAAATT TTCTTGAAAA ACTTTCAATA	6720
GAAAAACTTT TTGACAACCT GCAGCTAATA ATTCATGGT ATCTTTATTG CCAATCGTCT	6780
CTTGCTCTTC TACTAATTCA TAACCAGCTG TAATACTGGT ACAAATTCG CATCCTTTTG	6840
TGCCTCGTTG ACACACCCCT CGTTGACCAC TAATCGTAAT AATCTTCAAG CCATAAATTG	6900
GATAGACTTC TTGGTACTGT TTAATAAGGG CTTCGGCTAA TGTCGTTTGT CCACTATTCC	6960
GACCAGTAGA ACCAATTTGA ATCATGTTTG GTATCTCAAT TGCCATAGAG CTCCTCCGTC	7020
TGAATCGTTT TCTCAACTCT TGTTTGACA GTTGCTGGTT TTGTCATAAA GATCAATGGT	7080
ACGGCCATAA AAAGACTGCC ACCTACTGCA TTGCCTAAAA ATACCCATAA TGTATTCTGC	7140
AAATAATCCC ACCAACTGAT TGTGGTACTT TGTGAAAAA TTGCCGCTGG AATAATAAAT	7200
GCATTAGCGA CCACGTGTTG AAACCCACTT GGGGATTACA ATTCAAAATG AGAT	7254

## (2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7296 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

```

GGGGCCAATT TaGaAAATkG ACmATTGAAT AGAAAAGGaT GTGAACaAAT GGCaAAAGAA      60
ACAAATGTTA AgTTAGTCaC GGTGAGTGTT TTTGTGGCAA CATTTATGAC AGCCATTGAA      120
GGGACCATTG TGTCTACTGC GATGCCAACG ATTGTGCGCT CGTTACATGG CATGGAAATT      180
ATGAACTGGG TATTTTCAAT TTATTTATTA ACGAATGCGA TGTTAACACC GATTTATGGG      240
AAACTTGCGG ATAAAATTGG TCGTAAACCT GTCTTCATGA TTGGCATTAT CATTTTATT      300
TTGGGCTCCT CGTTGTGTGG CTTTGCTCAA GATATGTTGA CTTTAATTAT TGCCCGCGCA      360
ATTCAAGGTG TGGGGGCAGG CGCAATTTTA CCAGTTGCGT TAACGATTAT TGCCGATATG      420
TATACATTGG ACAAGCGAGC GAAAATTTTA GGTtAAACA GTGCCGCTG GGGAAATTGCT      480
AGTATTTTGG GTCCGTTAGC AGGTGGTTTT ATTGTAGATA CAGTCGGTTG GCATTGGATT      540
TTCTTCATTA ATGTTCTAT TGGACTTGTT TTATTGGGCT TGATTAGTAT TTTCTTAGTT      600
GAACCAAAGC GGGAAACGGAC CAAGATGCCA ATGGATATTT TGGGCAGCGT TACTTTGATG      660
GTAGTGCTGC TAACGTTATt GCTAGGTTTT CAAATGATTA GCGATAATGG TTTTACATTA      720
GTAACATTTG GTTGTTTAAG TTTGAGTG TGCTTTTTTG TAGCATTGTG GATGATAGAA      780
AAACGCGCGC AAGACCCAGT GATTGATTTG CATTTATTTA ATCAACCAAC GTTTGTTTTA      840
GTAAATCTTA TTGCAGCGCT TATTAGCGGT TTCTTAATGG GGATTGATGT CTACATTCCG      900
ATGTGGATGC AAGGTGTCTT AGGAAAAAGT GCAGGAATTG GTGGCTTAGT TTTAGCGCCC      960
TATGTCGTTA CTTTGATGG CTGGATCATT TATAGCAAGT AGTTTTATGG AAAAATATGC      1020
TATGAAAAAA GTCTTAACGA TTGGGTATC GATTCTATTA GTCGGGGCCA TCTTTTAGT      1080
GGTAATGCCA ATGGCCGTTT CGTTTTGGCT TTTCTTTGTA GTGTCTTCTG TCTTAGGAGT      1140
TGGTTTTGGG ATTACTATCA CAACGACTAC GGTGACAGCA CAAAGTACAG TGGAGCCTGA      1200
AAAAATGGGG GTTGCAACAT CGTTTAATAC GTTGGTGCGT ACAATTGGGC aGACTGTGAT      1260
GGTGTCAATT TTTGGTGTGA TTTTAAATGc AGGAATGTTT GCGAAATTGG AAGCGAGCGC      1320
GTTAAACGTC GATGCAGATG TCATGAATCA ATTAGTGAAT CCACATACTG CaAATTaATT      1380
CCAGCTGCGT TGTTAAACC ATTACGCGGT ATCCTCTATG CAGGTCTGCA TAATGTTTAC      1440
TTAGTCGGTG CGGGCTTAGT TGTTGTCGCT CyTTTGTTAA ATATTTTCGC AAAAGCGCAA      1500
CGAGCGAAGG TTTAGTGGTA CACTAATAGA AAGAGTATTG GAGGGAAAAG TAGATGCCT      1560
TTTTACAAGC AAATATTTAT TCmAATGTAT TGGAAATGGA AGTGAATGTG AACGTAATTT      1620

```

TACCACAAGA AACGGTAAAA AAAGTGGGCA CAAGTACCCA AGGAGCATTA ACAGATATTC	1680
CCGTGTTGTA TTTATTACAT GGGATGGGAG GGAATCACAG TGTCTGGGCT CGCAGAACCT	1740
CTATTGAGCG GTATGTCGCT GATTTGGGGA TTGCTGTCAT TATGCCGTCC ACGGATTTAG	1800
CTTGGTATAC GGACACACAA TACGATATGA ACTATTGGAC GTTTATCTCA GAAGAATTAC	1860
CGAAAAATTTG TCATCAACTC TTTCCACAGT TAAGTACAAA ACGAGAAAAA ACCTTTGCCG	1920
CTGGTTTGTC AATGGGTGGT TATGGCGCCT TGAAATTAGG ATTAGCGAAA CCAGAATCTT	1980
TTGCAGCCGT TGCTTCTTTG TCTGGGGCAG TCAGTTTATC ATCCACTAGT TTTGGGGAAT	2040
TATTAAAAGT CCGTAAGCGA AgcTATTGGG AAGGTATTTT TGGCCCGTTA GATCAAATTG	2100
AAGGTTCAT TCACGATCCG CTGTATTTAT TACAACAATT AGTAGAAAGC CAAACAGAAA	2160
TGCCCCAATT TTACCTTTGT TGTGGGGAGC AAGATATGTT ACTTTCTGCT AATCAACAGA	2220
TGGCTCAAGC GTTGAACAG GCACAGGCAA GTTACACTTT TGAACTGGA CCAGGAGAAC	2280
ATGACTGGGT GTTCTGGGAT GAATGGATTC AAAAAGCCTT GGCTTGGCTA CCAATACCAA	2340
AATAGAGAAA GTAGTTTCA ATATTTATTT TAAAAGTTAA ssCTTGGTCC GTCAGTCGAA	2400
ATGACTGACG GACCAAGGGc TTTTTTTATT TAGATGAAGA TGTGTTGTCT GCTTGCAACC	2460
CATCTAAGTA GGTTTTAACG GCTTGAATAA AGGCGGATTT TTGGTGGATG GAGTGTGAA	2520
ATTGAATGCT TTCCTCCATG TACGTTAAC TATTGTGCGA TACATATTGT TTCATGTGCA	2580
ACACTTGTTT TTTTAAATAG TCATCTCCGT TCAACACTTG CGAAACGAGC TGTTTTCTCT	2640
CCGTTGCGTC CATAATCATC CCAAATAAAA GATGTTGTAA AGACGTTACT GCTAAATGAG	2700
TCATTTCTGG TGAAAAACCT GCTTCTCTTA AAATGCCCAT CATTTGATTC AAATGGCGTA	2760
ATCGTTGGGG ATAGGCAGGG ACGGTTTGGA TTTCGATAGC TACGGCACAG GGAAATTGTT	2820
GATAGAGATC ATAATAGTTT TCCATGAAGG CTAGTAAGTC AGAATACCAT TCACCACATA	2880
ATGCAGGTTC TTGAAAGTGC TCTTCAATCG CTTAGCCAT ACTTTGAAGT AAGGCTTGCT	2940
TATTTTTAAA ATACmAtaA ATTGcGGGCG CCTGTACACC AAGCTGTTG GCGACTTTAC	3000
GCaTAGAAAG TTGtTCTAAA GTGGGACTTT TTTCTAACA AGAAAAATGCG GCTGCAATAA	3060
TCGTGTCTTT TGATAATTTT GGTTCCATGA TTTTAGCTCC TATTAAACGT TATCTTCGTA	3120
AGTTTATCAA AAAGAAGGTA CACTTTATGA TAGGAAAAAC TTGTGAAAAA TTAATTGTT	3180
CGCTTGATTT TTACCACTAA AAATGGTAAA TTATGAATCG GCCTTATTTT ACAGTGTTAA	3240
ATTGGTGAAG TCAATGAGCA AATTGATTTT TTTGTAATGG GGAAGAATTC GCAGACCAAC	3300
ACTGAAGAAG CAGGGAAC TA TTTAAGAAA GTAGGGATAA GATGGACCTT ATTATTCAAC	3360
ACGCCAAAAA ATACAAAGGC TCGGTAGTTA TTGCCTTGTT GGCGGTGATC GTAATGGTTG	3420
TCTCAGCACT TTGGCAACCA AAATTGTTAC AACAAGTGTT GGAAGCAATT ATGAATGACG	3480
ATAGCGACAA AATGAAAAAC TTAGGGATTC AATTAATTGC CATTGCAGGA TTAGGACTCG	3540
TTGCAGGGGT TATTAATACA ATTTTTTCTG CCAAAGTTGC CCAAGGTGTC AGTGCTGACA	3600

TTCGTGAAGC AACTTTCCTGC AAAATCCAAA CCTTTTCATT TGGCAATATC GAAAAATTTT	3660
CAGCAGGTAA CTTAGTTGTT CGTTTAACAA ATGACGTTAC TCAAATTCAA AATGTCATTA	3720
TGATTGCGTT GCAGACACTT TTTAGAATTC CTTTTTTATT TATCGGGAGT TTTATTTTAG	3780
CGATGCTAAC CTTGCCACAA CTATGGTGGG TGATTGTTGC ACTTGTCATC GCAGTTATTT	3840
TGATTTCTAT GCTGTCCTTT TCGCAAATGG GCAAGCATTt CATGATTATT CAAAACTTGA	3900
TTGATAAGAT TAATGGTATT GCAAAAGAAA ATTTGTTAGG CATTCGTGTC GTCAAATCAT	3960
TTGTTCAAGA AAAAAATCAA TTGTCACGTT TTACAAAAGT GaGTGAAGAA TTaAcCACCC	4020
ATAACTTAAT CGTAGGTAGC TTATTTGCTG TGATGATTCC AGCGTTTATG TTGGTGGCGA	4080
ACTTAGCTGT TGTAGGCTCA ATTTTCTTTG TGAGTAACTT AGTCAAGGAC GATCCAACTT	4140
TAATCGGTGG TGTCGCTTCC TTTATGAACT ATTTAATGCA AATCATGATG GCGATTATTA	4200
TCGGCGGAAT GATGATGATG ATGACTTCTC GTGCCGCTGT TTCCATTAAAG CGGATTAAAG	4260
AAGTGATGGA AACAGAACCA GATGTTACCT ACAAAAAAGT CCCAGAACAA GAGTTAATTG	4320
GTAGTGTTGA GTTTGACCAC GTATCTTTCC GTTATCCTGG GGATGAAGAA GATACCTTGA	4380
AAGACATCTC TTTCTCTATT CAGCCAGGTG AAATGATTGG GATTGTGGGC GCAACTGGCG	4440
CTGGTAAATC TACATTGGCA CAATTAATTC CACGTTTGTT TGATCCGACA GAAGGAAAGA	4500
TTGAAGTTGG TGGTGTGAT TTACGAGAAG TGAACGAACA CAGTTTACGA AAAACAGTTT	4560
CCTTTGTGTT GCAAAAAGCA ATCCTGTTCT CTGGAACGAT TGCACAAAAT CTACGCCATG	4620
GTAAACGTGA CGCTTCTGAA GCGGACATGG AACGGGCTTC AGGCATCGCC CAAGCCAAGG	4680
AATTTATTGA AAAATTAGCA GAAGGTTATG ATGCACCCGT CGAAGAACGG AGTAATAACT	4740
TTTCTGGTGG ACAAAAACAA CGGTTATCAA TTACAAGAGG TGTGATTGGT GAACCTAAAA	4800
TTTTGATTTT AGATGATAGC ACCAGCGCGT TGGaTGCTCG TTCGGAGCGG TTAGTCCGAG	4860
AAGCGTTGGA TAAAGAGTTA AAAGAGACTA CCACGATCGT TATTGCCCAA AAAATTTCTT	4920
CTGTTGTCCA CGCAGATCGT ATTTTAGTAT TGGATAATGG ACGTTTAGTC GGGGAAGGCA	4980
CGCATGAAGA ATTAGCAGCT ACGAACCCCG TGTATCAAGA AATTTACGAA ACTCAAAAAG	5040
GAAAGGAGGA AGCTTAATGA CTGATTTAAT AAAAGCTAGT AAATTTTCT ATCATTATTT	5100
GAAACGTTAC AAAGTTTCGT TCCTCTTTAT TTTCTTAGCA ATTTTCGCAG CGACTTATTT	5160
ACAAGTCAAA GCGCCGCAAT TCGTTGGGGA AGCTATTCAG GAATTAGCGA AATATGCGGT	5220
TAATGTGATG CAAGGAAAAG ACGATAAAAG TCGGTTTCGT TCTGTCATTT GGAAACTACT	5280
CATTTTTTAT GTCTTAACTA GTGCCGCTAG TTTCAATTAT AGTATTCTCT TTACACAAGT	5340
CGTGGGGAAA TCGACGAACC GCATGCGGAT TGGTTTGTTT AACAAATTAG AAAAAATTGAC	5400
GATTCGTTTC TTTGATTCTC ATCAAGATGG TGAAATTTTA AGTCGTTTTA CTAGTGACTT	5460
AGACAACATC CAAATAGCT TAAACCAAGC GTTGCTACAA GTATTAACCA ATATTGCCTT	5520
ATTGGTTGGT GTCTTAATCA TGATGTTCCG TCAAAATGTG GAACTGGCAT GGGCCACAAT	5580

TGCTTCTACG CCGATTGCGA TTTTAATTGC GGTCTTTGTG ATTAGCAAGG CGCGCAAATA	5640
TGTCGATTTA CAGCAAGATG AAGTGGGTAA ATTAAATGGC TATATGGATG AAAAAATTAG	5700
TGGGCAACGT GTGATTATCA CTAATGGCTT ACAAGAAGAA ACCATTGACG GCTTTTTAGA	5760
GCAAAATGAA AAAGTTCGTG CCGCTACGTA TAAAGGTCAA GTGTATTTCAG GATTACTTTT	5820
CCCAATGATG CAAGGATGT CATTAGTCAA TACGGCGATT GTTATTTTCT TTGGTGGTTG	5880
GTtAGCAATC AATGGCTCTG TTGATCGTGC CGCTGCGCTA GGTTTAGTCG TTATGTTTGT	5940
CCAATATTCT CAACAATATT ATCAACCATT AATGCAAATT TCTTCAGGTT ATAGCATGAT	6000
TCAATTAGCT GTTACAGGCG CTCGTCGTTT AAATGAGATG TTTGATGAAC CAGATGAAAT	6060
TCGTCCCAGAA AATGGCGAAA AATTAGAAGa GATTAAACAA GCTGTGGCTT TAAATCAtGT	6120
TGTTTTTGGC TATAATCCAG AAACGCCTGT GTTGAAAGAT GTTCTATTc ATGTGGATAA	6180
AGGTGAAATG GTCGCTTTAG TTGGACCGAC AGGCTCAGGA AAAACAACGA TTATGAATTT	6240
AATGAATCGC TTTTATGATG TGAATGAAGG CGCCGTGACG TTTGATGGTG TGGACATTcG	6300
TGAAATGGAT TTAGATAGCT TCGGTTcACA TGtAGGCATT GTTTTGCAAG AGTCCGTtCT	6360
ATTTTCAGGA ACAATTCGTG AGAATATTGC TTTTGGAAG CCAGAGGCAA CGGATGAAGA	6420
AATTGTTCAA GCAGCGAAAC AAGCGAATAT CCATGAATTT ATTGTAAATC TTGAGCAAGG	6480
CTACGATACT GAGATTACGG AAGAGAATAA CCTTTTCAGT ACGGGGCAAA AGCAATTAGT	6540
AAGTATTGCA CGAACGATTA TTACCAATCC AGAATTATTA ATTTTGGATG AAGCAACAAG	6600
TAATGTTGAT ACAGTAACAG AAGCTAAAT CAAAAAGCG ATGGACGAAG CAATTAAAGG	6660
GCGCACCAGT TTTGTGATTG CCCATCGTTT GAAACGATT CTAAATGCGG ATCGGATTAT	6720
TGTTTTACGA GACGGTGAAG TCATTGAAGa AGGCAACCAC CATGAACTAG TTGAACAAGA	6780
TGGATTCTAT GCGGAACTTT ATAAAAtCAA TTTGTTTTTG AATAAGTCGA AAAAGAAGCA	6840
AAATGAGTCA TTCTCATTtG CTtCTTTTTT TGCTATGCTA AATAGAAAAA GGAGAAATGT	6900
CAATGGAAAA AATGAGGAAT ATACGGCAAT TACGGCGCAC GTTTTTGTTG CTAAATACTT	6960
TATTTGTACT GAGTGATTc AGAGAGAATC GATGGACAAA ATCGGCACTG CTTTTTTTGG	7020
TTGTTGGCTA ATAGTTTTTT ATATGGACTG CAGTAGTTGA TTTATCTTGG GGGCGAAAAA	7080
GGCTGGCTGA TACATTTGAA GCGGTGGTAT TATTGCGCGG TTCTTCTCCA AATAGCAGGC	7140
ACATTGATTG TATATTTTTT GTTTTTTGTA TAAAGGTCTC TCGTTTCTTG CCGGAGAATA	7200
TGCTAAAATA AAAAGCAGCA GGGGGCGAGA AAACCATGGA AAAGCCnAAT nAACGCCnAG	7260
CTGGATTTAG AACGTACACG GGCCAATTAA TTTTAA	7296

## (2) INFORMATION FOR SEQ ID NO: 60:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9927 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

GATTAAATGT CCTGCCTCT GTTTTGGAAA ATCCTAAATC AAGCCTTTCA CACCTATATC	60
TTTTCGGTAG TACATTCCAG GAAATGAGTA ATTATCCAAT GTTTGATAAA CTTTTTTGCG	120
GGCTTCTTCA AGAGTCGAAG CAGTTGCTTC TGCTAAAAAG ATTCGTCCGC CAGCCGAAAC	180
CAGAGCCTGT TCTGAACGAG TTACCCCTGC ATAATAGATG GTTGCACCAG ACATTTCTGG	240
CAATGGTAGA TTTGTCTTAT ATTGATTAGG ATAGCCTTCT GCTGCTACAA CAACACCAAC	300
ACTCGCTAAC TCATTCCATG CAATCGTTGG TTGCTCTTTC ACCAGCAATG AAGAAATTAT	360
TTCTGAAAGG GGACTCTCTA AACGTGCCAA CACAATTGT GTTCTGGAT CCCCAAATCG	420
AGCATTAAAT TCAATAAAGT TGGGGCCAGT TGCTGTTTGG aTTAAACCAG CATATAAAAT	480
CCCTTGAAAC GGACGTTTTT CAGCCACCAT TCCTTTAGCA GTCGGTTTCA GGACTTGGCT	540
AATGGCTTGC TGAACCGCTT CTTCTGGGAC CTGAGGCACT GGTGTATAAG CGCCCATCCC	600
CCCTGTATTC GGACCAGCAT CCCCATCGAA AATCCGCTTA TGATCTTGAG CAATCACCAT	660
CGGATATACC TTCTCATTAG AAACAAAAGC CATTAAAGGAA AATTCTTCTC CTTGCAAAAA	720
TTCTTCGATT ACAACCGTTT GACCGCTTGC TCCAAATTGT TGGTCCTCCA TCATCGCAAC	780
TAGCGCCCGT TTTGCGGCTG CTTTTGTTTC GGCAACTACT ACTCCTTTTC CAGCTGCCAA	840
ACCATCTGCC TTAATAACAA TTGGTACTCC TTGCTGTTCA ATATATTCTA AGGCAGGTGC	900
CAAACGTCGA AAGACCCGAT ATTCCGCCGT TGAATCTGG TATTTTTGCA TAAAATCTTT	960
CGCAAAACTT TTTGAACCTT CAATTATTGC AGCTGCTTTC GTCGGGCCAA AAATCAGTAA	1020
TCCCGCTGCT TGAAAGGCAT CTACAATTCC TTCAATTAAT GGCCTTTCCG GTCCGACGAA	1080
TGTCCAATCA ATCTGATTAT CCTTCGCAAA TTGAATTAGT TCCGTTTGCT GCTCTTCACC	1140
GATAGGTACA GTCCGAATAT TATCACCCTT CATAACAGGA TTACCTGGTG CACAATAAAC	1200
AGTCTTGACT TTGTCTTCTT GCTGTAATTT TTTGGCGATT GCATGTTTAC GACCACCACG	1260
CCCAATCACT AATATATTCA TCGGTTTCTC CTTTAAAAAT TAATGTTTGA AATGACGAGT	1320
CCCAGTAAAA AtCATCGCAA TACCATAATG ATTCGCTAGA TCGATTGATG CTTGGTCTTT	1380
AATGCTCCCG CCTGGCTGAA TAATTGCTTG AATGCCaTGT TGTGCAGCAT ATTCCACGCT	1440
ATCTGCCATG GGGAAAAAGG CATCACTTGC TAAGACGGCC CTTGTAATT TTCCAGCTGC	1500
TTGGGCTTGA TCAACTGCAA TTTTACTGA ACCAATTCGA TTCATTTGAC CAGCGCCAAT	1560
CCCCACCGTT TGGCTTTCAT TGGCTAAAAC AATAGCGTTT GATTTTACAT GCTTGACTGC	1620
TTTCCAAGCA AAGTTCAATG CGTCTCTTTC GGCTGGTGTG GGTGGCGCT CTGTAACAAC	1680
TTGCCATGCT TGGTCATTTT CAACAAGTGT ATCCTGTTCT TGAATTAATA ACCCGCCCAT	1740
AACAGAAACT AATTCACCG CTTTTTTTTT TTGAGCATGA AAATCTAACG TTAACAAACG	1800
AAGATTCTTT TTCTGAGATA ACAGTGCCAA AGCTTCTTCC TCAAACTAG GCGCAATAAT	1860
TATCTCTAAA AAGATTTGAT GCATTCAGC CGCAGTCGCT TGGTCCACTG GACGATTCAA	1920



AACAACAATG CCACCAAAA TTEAACTGG ATCTGCTTCA AATGCCTGAC GATAGGCTGC	1980
CAAAATCGTT TTTCTGTTC CAATTCCACA TGGGTTTATA TGTTTAACAG CGACGACCGT	2040
AGGTTCTGTA AACTCSCTAG CATTCCGGAG TGCTGCATCA GCATCTCGAA TATTGTTATA	2100
TGACAGTTCT TTGCCATGAA GTTGTGAGC GGAAGCAATG GATAATGATA CAGGTACAAC	2160
GGATTGATAA AAGGTGGCTT GTTGATGGCT ATTTTCGCCA TAACGCAATG TTTGCTTACG	2220
TTCATAAGTA AGCGTCTGTT TCTCTGGTTC AGTTTCACCC ACTTGCCTCG TTAAATAGTC	2280
AGCGATAAGC GCGTCATAAG CAGCGGTATG ACGAAAGACT TTTGCTGCTA ATCTTTTTCT	2340
AGTGGCTAAC TGGCTAGTTC CAGTGCTGCT AATTTCACTT AATACCTGTT CATAATCACT	2400
AGGATCAACC ACTGCTGTTA CGTCTTGATA ATTTTTTGCC GCTGAACGTA ACATGCTAGG	2460
ACCGCCAATA TCAATCATT CAATCGCTTC TGCTTCAGAG ATCGCTGGTT TTAAATTTGT	2520
TTCTTTAAAT GGATAAGAT TCACTACTAC AAGGTCaATA AAGTGAATGC CTTGTTCAGT	2580
CAGTGCTTTT TGATGCGCTT GGTTCCTTCT TTTTGCCAAA AGTCCGCCAT GGATATTCGG	2640
ATGTAACGTT TTGACACGGC CATCTAACAT TTCTGGAAAG CCAGTGATGT CGTCAATAGC	2700
AAGTGTCGGT ACCCCCGCTT CTTCCAAGAC AGTCCTTGTT CCCCAGTGG AGATAATTC	2760
AAAGCCATT GCAACTAAGC CTGCTGCAAA TGTTGCTACT CCTGTTTTGT CTGAAACGCT	2820
AATCAATGCT CTTTTGGTCA TTTCATTGTC TCCTTTACAA TTTGTGAAAT TATTTTTGGG	2880
TACCATTCTG GCTCAAGAGC ATGAATTTTC TCTGCTAATG TGTCCACTGT ATCTTCGGTA	2940
TCAATCTTGG TGTTTSTTTG GAAAATAATC GGTCTGTAT CAACCCAGTA GTCAACaTAA	3000
TGGATAGTGA TTCCTGTTAT TTTCACTCCA TAATGAAATG CTTCTTCAAT TCCgTGAAGA	3060
CCTGGAAAAC TAGGTAAACAG AGAAGGATGA ATATTGACGA TTCTCTTTGG ATAAGCCTCC	3120
AGTAAAGTCT TGCCGATAAT TCTCAAATAA CCAGCTAACA CGATTAAATC GATCTGATGC	3180
TCTTTTAAAT GTTCAATAC TTGTTCTCTA TATTGCCAC GAGATGAAAA ATCAGATGGC	3240
GAAAAGCAAA CAACGGGAAT CTGCGTTTT TGGGCACGGG TCAATACATA GGCTTCAGGC	3300
TGGTCACAAA AGACTAAGCT CAAATGCCCT GCGATTTTT TTTGGGAAAA AGCTGCAGCA	3360
ATCGCCTCAA AATTGCTACC ATTGCCAGAT GCAAAAACAG CTATCTTCAT AAGGCTTCCT	3420
TAAAAATAAC GGCTTTCGTT TTCCTAGGTA CCAAACGACC AATTTCAAAA GCTTCTGGCA	3480
AAAGCTCCTG AACTTTTTCA ACATGCTCTG GACTCACCGC TAAAATCATC CCTAAGCCCA	3540
TATTGAAAAT TTCATACATT TCTTCAGGTG GAATTTCAAC GTATTTTGA ATCAATTTAA	3600
AAATTGGTAG CACTGGCCAA CTATTTAATT GAATCTCTGC AGCTAATGCG GAAGAAAACA	3660
TTCTTGGTAG ATTTTCTAAA AACCCGCCGC CAGTAATATG CGCAGCCCCA TGCACGAGTC	3720
CTGCTTTTAA CAATGGCAAT AATTCCTTCA CATAAATTTT TGTTGGTGTT AATAATTCTT	3780
CCCCTAACGT TTTGGGTGCC AATTCTGGAA GCTTGTCCGT TGTTTTAAAC GAATGTTTCT	3840
TAAAAAAAT ATTTCTTACT AAGAGTAAC CATTAGAATG AAGACCACTT GAAGGAAGAC	3900

CAATCAAGAA ATCGCCTTCT TTTACGTTAC TTGGAGTCAA GAGTTGACTT TTCTCAGCAA	3960
TACCCACGGC AAAACCAGCC ACGTCATAAG CATCTGCTTC ATACATATCA GGCATTTTCAG	4020
CTGTTTCGCC CCCAATTAAG GCTGCATTTC CCTCACAACA GCCAGCCGCA ACACCCGCAA	4080
CAATTGCTTC AACTTTAGCT GGATTAACGG TCCCTAACGC TAAATAATCT AAAAAATATA	4140
ACGGTTCGGC GCCTTGTCG ACGACATCAT TGACACACAT TGCTACACAA TCAATTCCGA	4200
TTGTCTCGTG TTTCTGTTCT TCAATTGCCA AGAGTAACTT TGTGCCTACG CCATCCGTTT	4260
CTGATACCAA AACTGGTTCT TTAAGCTTGT AACTACTTAA ATCAAAACAA CCACCAAAC	4320
CACCAAGCAT GCCCAACGTT CCCGTCCGTT GTGTTTTTTG GCTATGCTTT TGAATTCGTT	4380
CTACTACTTC GTATCCCGCC TCAACATTTA CACCAGCTTT TGAATAGGCA TTTTCCATGT	4440
CAATTCCTCC TAGTTTCTTT TTTCAAGGAT GCTTGATATT GTTCTTCATA ATCATAACAA	4500
GGCGTTGGAT AGTCTCCATT AAAGTAAGCC ATACACAGCC CAGAATAAGG AGCATCATAA	4560
TTTAATTGAA TCCCATCAAT CAGCCCTGCT TCACTCAAAA AAGCCAGCGA ATCTGCGCCA	4620
ATCTGCTCTT TGATTTCTTC AATTGAATAG TTTGCCGCAA TTAGTTCGTC CCGGTCTGT	4680
ATATCAATGC CATAAAAACA CGGATATTTT AATGGTGGTG AAGCAATTCT GACATGAACT	4740
TCTTGCGCTT CTGCTTCTTT TAATAAATGA ATAATGCGCC GAATAGTGGT CCCGCGAACA	4800
ATCGAATCAT CTACTAAAAT GACTTTTTTTT CCTTTGACCA CGCCACGAAC TGCCGATAGT	4860
TTCATACGGA CACCTTGTTT TCTTAATTCT TGTGTCGGCT GGATGAATGT TCGAGCGACA	4920
TATTGATTTT TACTAAACC TAGTTCGTAA GGTATCTGTG CTTCTTCTGC ATAGCCGCTA	4980
GCTGCAGACA ATGACGAGTT CGGTACACCA ATTACCATAT CAGCTGCGAC TGGACTTTCG	5040
ATTGCCAAGT TCTTCCCAT TCTTTTACGA GCTGTATGCA CATTAACTCC GGCAATATTT	5100
GAATCTGGTC GTGCAAATA AATATATTCC ATTGCACAAA TTGCCGGTTG GACTTCCGTA	5160
GTGAAAACCT CTATTTTCAA GCCTTCATCA GAAATGATGA CTACTTCACC AGGCGCCACA	5220
TCTTGAATAA AGCGAGCCCC AACTGTTTCT AATGCACACG TTTCGGATGC CACAACGTAC	5280
GCCCCATTGT TCATTTGTCC GATGGAGAGA GGGCGAAAGC CGTTGGGATC CAGTGCAGCA	5340
ATCATCATCG TTTCTGTAA TAAAAGATAA GCAAAGCCAC CCTTAACCAA ATTCAAGCTC	5400
TCTTTTAATT TATCGAGAAA CGAGGCAGCA GTGCTTTTTC GAATCAAGTG CATTAAAATT	5460
TCAGTGTCTG AATTGGAATG AAAAATAGCG CCTTCTCTTT CTAGTTGTTT TCGTAAACGT	5520
TTGGCATTGG TCAATTGCC ATTGTGTGCT AaCCCAaTtG ACTATCATAA AATTTAAATA	5580
AAAATGGCTG AATATTATCC ACGCTACCCG TACCAGCGGT TGCATAGCGT ACATGACCAA	5640
TAGCCGCCGT TCCTTCTAAC GCATTCAAGT GTTCTTGCTT TTGGAATACT TCTGCTAATA	5700
AACCTAAATC ACGGTGGCCA TCTAATTTC CAGAATTGTT CGCCACAATT CCCGCGCCTT	5760
CTTGGCCTCG ATGCTGCAAG CTGTGTAAGC CAAAATAGGT CACTTGTGCC GCATCAGGAT	5820
GACCCCAAAT ACCGAAAATA CCACACTCTT CATTTAAGCT TTTTGCTTCA TAAGACATGG	5880

AATCGCATCC TCCATAAGG CTTTGGCTTC TTTTGTGAG CAATTGATCA CATCATCTAA	5940
CGCATGAATG ACCAAACCAG TTTCAGTCAC TTTCCAATA TGTTGAGCTT TTCTTCCGAT	6000
AAGTGTTC AATGCTTCTT GATGCTGGGG ACTAACTGAA AGAATAAAAC GGGATTGCGT	6060
CTCCGCAAAT AGATATTCTT TTTTAAATGG CAGTGTACC TGTAACCCCA GTTCGTTAGC	6120
AAAAGCCGAC TCGGCTAAAG CAACGGCAAC GCCACCTTCA GCGCAATCAT GCGCGCTAGC	6180
AACTAGGCCT ACTTGAATTG CTTTTAAAC TAATTCTTGG TTGCTTTTT CCTCTTTAA	6240
ATCAAATGAG CGTAATTGCC CTTCAATTCG ACCCAGTTGT ATTTTTTGAA TTTCTGAGCC	6300
GTTAAATCT GCGTGGGTTT CACCGATGAG ATAAATTAAG TCATCAACTT GTTTAAACGC	6360
TTGTGTGGTA ATTTGTGAAA CATCTTCAAT CACTCCCACC ATGCCAATCA TTGGCGTTGG	6420
ATAAATGGCT TGACCATCTG TTTCATTGTA AAGAGAGACG TTCCCTGAAA TCACGGGTGT	6480
TCCTAGCTGA CGGCAGGCGG CGGCGATTCC ATCCGCCGAC GTCCATAATT CCCAAAACT	6540
TTCTGGTTTA TCTGGGGAAC CATAATTCAA GCAATCAGTG ATAGCCAAAG GTTGTGCTCC	6600
ACTCGCTACA ATATTACGGG CTGCTTCTGC AACCGCAATT TGCCACCAA TTTCAGGATT	6660
TAAATACAAA TAACGGCTAT TACAATCTGT CGTCATCGCT AAGGCTTTAT TGGTGCCCCG	6720
CACTCTTAAA ACAGCTGCAT CACTTCTGG TTGCACCACT GTATTCGTTT GTACTTGACT	6780
ATCATAGGTT TCGTAAACTG ATTTTTTCGA TGCAATCGTT GGCTGTTGCA ATAGGCGTTT	6840
TAATAATTCC GTACTATTCTG TAATTCTTGG AATGAAGGGC GCTAATTTTT TAAAAGCACG	6900
CATCCGCTCT GGTCTTGAT AGGCTTCTG GTAGACAGGG GCTTCTCTG CCAAAGCATC	6960
TACTGGTACA TCTGCGACTA TTTTGCCACC GTGAAAGACA CGATACTGTC CATCATCTGT	7020
TACGGAGCCA ATCGTACCG CTTCTAAGTC ATACGTTTGA AACAAATCAC AAACCTGCTG	7080
TTCTGCTCCT TTTTCCACAC AGATTAACAT CCGTTCTTGG GACTCAGACA AGAGCATTTT	7140
ATAGGGGCTC ATCTGCGTCT CTCTCTGCGG AACTTTATCT AGCTCTAGAA TCAAGCCTGA	7200
AcCTGCTTTT GATGCCATTT CAGCACTTGA AGAAACCAAG CCAGCCGCC CCATATCTTG	7260
GATGCCGATT AAAATATCTT GATGATTTTT GATTAAATCA AGACAAGCTT CCAATAATAA	7320
TTTTTCCATA AACGGATCAC CTAATTGTAC CGCAGATCGT TGCTGTTCTT CACCTTCACT	7380
AAATTCTTCT GAAGCAAAGG TCGCACCATG AATTCCATCT CGTCCTGTTT TCGCACCTAC	7440
GTAAAGAATG CTATTTCCGA CGCCTTTTGC TTGGCCTTTT TGAATCTCAT CATGACGAAT	7500
CAATCCACA CACATAGCAT TAACAAGTGG GTTTCCTTGA TAACATGGCT CAAAAGCAAT	7560
TTCGCCGCCA ACGGTTGGAA TACCAATACA ATTTCCATAA CCGCTAATCC CCGCCACAAC	7620
TTCTTGAAAC AAATAACGTG TTCGTGCATC CGTCAGTTG CCAAATCTTA GCGAATCCAG	7680
TAAAGCAATC GGCCGCGCCC CCATACTGAA AATATCGCGA ATAATTCCAC CAACGCCTGT	7740
AGCTGCTCCT TCGTAAGGTT CTAATGCAGA AGGATGATTA TGGCTCTCTG CTTTAAAAAC	7800
AACAGCTAAT CCGTCACCAA TATCGACAAT CCCAGCTCCT TCACCAGGGC CTTGAAGCAC	7860

TTGTGGCCCT TCCGTCGGAA ACTTTTTTAA AACGGGCTTG GAATTTTTGT AAGAGCAATG	7920
TTCACCTCCAC ATCACGGAGA ATAAGCCAGT TTCTGTATAA TTAGGTTCTC TTtGCAAAAT	7980
GGTCTCAATC TGATGATATT CTTCATCTGT TAGTCCCCAT TCTGCATACA ACCGTGTTTC	8040
CTTGATTCTT TTTGGTGTG GTTCATTCAAT CATTGGCTTA ATTCTTTTCC AAAGTTTTTA	8100
ATGATTGAAG CAAAAAGCT GCGTCCATCT GTTGAACCTA ACAATTCTTC CATCGCTCGT	8160
TCCGGATGAG GCATCATTCC TAGAACATTG CCCTGTTTCT TACAAATACC AGCAATATTC	8220
GCAAGACTAC CGTTCACCTT TTCTTGATAC GTAAAAACAA TTTGGTTCTT TTCTTTCAGC	8280
TCCTTCAACG TTTTTTCATC ACAATAGTAA TTCCCTTCGC CATGCGCTAC TGGTAAATGT	8340
ATCGAAACAT TTTCAGGATA CTCGCTTGTA AACTTTGTGT GCGTATTAC .TTTCAATGGG	8400
ACACTTTTAC AAATAAACG AAGCGAATCA TTTCTTAACA ACGTGCCAGG AAGTAAGCCA	8460
ATTTCTGTTA AAATTTGAAA GCCATTACAA GTACCAAAAA CCATTTTGCC TTCTGTAGCA	8520
AAGCGAATCA CTTCAATTGAT AATTGGCGAA AACCGAGCAA TTGCACCACA ACGTAAGTAA	8580
TCGCCATAAG AAAAACCACC AGGCAACAAT ACGCCATCAA ATTTTGCTAA AGAAGTGGCA	8640
TCGTGTCGGA CAAATTCTGC ATCGATGCCC ATCACYTCTT TGACAGCCCA CAATAAGTCT	8700
GCATCACAAT TGGATCCTGG AAAAACAATC ACTGCCAATT TCATTAAAT TTCCTCCAA	8760
GATAAAATTT CAAAACGATA GGATTCCATA TTTACATTGG CTAATAGTTG ATCACAAATT	8820
TGTTCAATTT TCTCTTCAAT TGCTTCTGTC TGCTGGACGT GCATTTCAAA ATACTTACCG	8880
ACACGGATTT CGGCTACTTC TGTA AACCA AGCCGATGGA CTGCTTCCTT AACCGCTTCA	8940
CCTTGGGGAT CCAAATTTGA TTCTTTATAC GTAACATAGA CTTTGACAAA ATACaTACGC	9000
GCTTCCTCCT CAATTAGCTG TTAGTAAACG TTGTAAACT TCTTCATATA CAGGAATAT	9060
TTCTCCTAAT TCACGGCGAT AAACATCTTT ATCCAGATGT TCGTTTGTGT TTAAATCCCA	9120
TAGACGACAG GTGTCTGGAG AAATTCATC TGCCAGTAAC AATTGATTGG CCTTCGTCGG	9180
ACCAATCTCG ATTTTAAAT CGATCAAACA AATATTGAGG CGTTGAAAGA GTTGACTAAG	9240
CGCCTGATTA ATTTGTAAAG CCAATGCTTT AATCTCCACA ATTTAGCAG GAGTGGCTAC	9300
TTTAGAAAT TGAATATGGT CTTCTTAAT AAAAGGATCA TCCAGACGAT CTTCTTTGTA	9360
ATAAAATTCA ATAATGGGCG TCACTAATTT AGTGCCCTCT TCAATCGCTA AACGCTTTGA	9420
AAAACCTCCT GCCGCATAAT TACGTACGAC AACTTCCAAA GGAATCaTTT CTACGACCTG	9480
AATTAATTGT TCGTGTTTCA ATACTTTTTT GATAAAATGA TTGGGAATCT TCTGTTGCTG	9540
TAAATGTTCA AAAATTAGAC TAGTGATTG ATTATTTAAA GCGCCTTTCC CAAGAACTTT	9600
GTCTTTTTTT TGACCGTTTA AAGCTGTTGC TTGATCAAGA TATTCCACAA AAAGAACCGC	9660
AGCATTTTCT GTTTGATACA ACTTCTTCGC TTTTCTGTA TACACAAGTG CTTTCTTTTC	9720
CATTCTTCC CTCTTCTCGT TAGCTTCATG AACCTATTTC TTTCTAACTc AGCTGATTCA	9780
CATCaGACaA AAACCTTACAT TAGTTACAAA AAAACaATAA TCGTTCGTGT TTGTCGAATG	9840

TAAAAATAAA AAAGTTGCTT GGAGCAAGCA TTTATTAAGT TACGTTCCAA ATACCAACAG 9900

TCTCAATTTC CTTTAACGTT TCTGATA 9927

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7678 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TGTCTGCT	TCAGTAAGGG	CGTTTTTGGC	TGAAGCAAGA	TATGAGGTGG	ATGCCTGCAC	60
AGATGGAAAC	GAAGCACACA	CCAAGTTCTA	TGAAAACACC	TATCAACTGG	TTATTCTTGA	120
TATTATGCTG	CCCGGTATGA	ATGGGCATGA	ACTTCTACGT	GAATTTCTGGG	CGCAAAATGA	180
TACCCCCATT	CTGATGATGA	CAGCCCTGTC	GGATGACGAA	AACCAAATCC	GGGCGTTTGA	240
TGCAGAGGCA	GACGACTATG	TAACAAAGCC	ATTCAAGATG	CGGATTTTAC	TAAAGCGGGT	300
GGAAGCCCTG	TTACGGCGCA	CGGTGCGCTG	GCAAAGGAAT	TTCGTGTGGG	CAGGCTGACA	360
CTTCTGCCGG	AGGATTTTGA	GGTACTTTGT	GACGGTACGG	AGCTGCCCCCT	GACACGAAAA	420
GAATTTGAAA	TCCTTTTGCT	GCTGGTGCAG	AACAAAGGCA	GAACCTTAAC	CCATGAAATC	480
ATTTTGTCCC	GCATATGGGG	ATATGACTTT	GACGGTGATG	GCAGCACAGT	CCACACTCAT	540
ATCAAAAATC	TGCGGGCGAA	CTGCCGGAAA	ATATCATCAA	AACCATCCGC	GGTGTAGGTT	600
ACCGATTGGA	GGAATCATT	TAATGGAAAG	AAAAGGGATT	TTCATTAAGG	TTTTTTCCTA	660
TACGATCATT	GTCTGTTC	TGCTTGTCTG	TGTAACGGCA	ACACTGTTTG	CACAGCAATT	720
TGTGTCTTAT	TTCAGAGCGA	TGGAAGCACA	GCAAACAGTA	AAATCCTATC	AGCCATTGGT	780
GGAAGTGAAT	CAGAATAGCG	ATAGGCTTGA	TATGCAAGAG	GTGGCAGGGC	TGTTTCACTA	840
CAATAACCAA	TCCTTTGAGT	TTTATATTGA	AGATAAAGAG	GGAAGCGTAC	TCTATGCCAC	900
ACCGAATGCC	GATACATCAA	ATAGTGTTAG	GCCCAGCTTT	CTTTATGTGG	TACATAGAGA	960
TGATAATATT	TCGATTGTTG	CTCAAAGCAA	GGCAGGTGTG	GGATTGCTTT	ATCAAGGGCT	1020
GACAATTCGG	GGAATTGTTA	TGATTGCGAT	AATGGTTGTA	TTCAGCCTTT	TATGCGCGTA	1080
TATCTTTGCG	CGGCAAATGA	CAACGCCGAT	CAAAGCCTTA	GCGGACAGTG	CGAATAAAAT	1140
GGCAAACCTG	AAAGAAGTAC	CGCCGCCGCT	GGAGCGAAAG	GATGAGCTTG	GCGCACTGGC	1200
TCACGACATG	CATTCCATGT	ATATCAGGCT	GAAAGAAACC	ATCGCAAGGC	TGGAGGATGA	1260
AATCGCAAGG	GAACATGAGT	TGGAGGAAAC	ACAGCGATAT	TTCTTTGCGG	CAGCCTcTCA	1320
TGAGTTAAAA	ACGCCCATCG	CGGCTGTAAg	CGTtcTGTTG	GAGGGAATGC	TTGAAAATAT	1380
CGGTGACTAC	AAAGACCATT	CTAAGTATCT	GCGCGAATGC	ATcAAAATGA	TGGACAGGCA	1440
GGGCAAACCC	ATTTCCGAAA	TACTGGAGCT	TGTCAGCCTG	AACGATGGGA	GAATCGTACC	1500

CATAGCCGAA CCGCTGGACA TAGGGCGCAC GGTGCGCAG CTGCTACCCG ATTTTCAAAC	1560
CTTGCCAGAG GCAAACAACC AGCGGTTCGT CACAGATATT CCAGCCGGAC AAATTGTCCT	1620
GTCCGATCCG AAGCTGATCC AAAAGGCGCT ATCCAATGTC ATATTGAATG CGGTCAGAA	1680
CACGCCCCAG GGAGGTGAGG TACGGATATG GAGTGAGCCT GGGGCTGAAA AATACCGTCT	1740
TTCCGTTTTG AACATGGGCG TTCACATTGA TGATACTGCA CTTTCAAAGC TGTTTCATCCC	1800
ATTCTATCGC ATTGATCAGG CGCGAAgcAG caAAAAGTGG GCGAAgcGGT TTGGGGCTTG	1860
CCATCGTACA AAAAACGCTG GATGCCATGA GCCTCCAATA TGCCTGGAA AACACCTCAG	1920
ATGGCGTTTT GTTCTGGCTG GATTTACCGC CCACATCAAC ACTATAAATA TTTAAACTT	1980
AAATGATTTT GACCGACAGG TATAACCCTG CCGGTCTTTT TGTTTTTCGC CGCTACAGGA	2040
AAACTACAGA TTGACTACAG GGAAAGTACA GATACGCTTG CCATAATAAC AATCGTACCA	2100
GCCACAAATC GTAGTTTTAT TGCAAAGGAG GCATTCAATC AAATGGAAAA AAGCAACTAT	2160
CATTCCAATG TGAATCATCA CAAACGGCAT ATGAAACAAT CTGGGGAAAA ACGGGCTTTT	2220
CTATGGGCGT TCATTATCTC GTTCACAGTC TGCACGCTGT TTTTGGGGTG GAGATTGGTT	2280
TCCGTATTGG AGGCAACACA GCTACCGCCC ATCCCTGCAA CTCATACAGG CAGCGGGACT	2340
GGTGTAGCGG AGAATCCAGA GGAAACACT CTTGCCACCG CCAAAGAACA GGGAGATGAA	2400
CAGGAATGGA GCCTGATTTT AGTGAACAGG CAGAACCCCA TCCCCGCCA GTACGATGTG	2460
GAACTTGAGC AGCTGTCAA TGGTGAGCGG ATAGACATTC GGATTCTCC CTACCTCCAG	2520
GATTTGTTTG ATGCCGCAAG AGCTGATGGA GTTTACCGA TTGTCGCATC CGGATACCGG	2580
ACAACAGAAA AACAGCAAGA AATCATGGAT GAAAAAGTCG CCGAATACAA GGCGAAAGGC	2640
TACACCTCTG CACAGGCTAA AGCGGAAGCA GAACTTGGG TGGCCGTGCC GGGAACAAGC	2700
GAGCATCAGC TTGGTCTTGC TGTGGATATC AATGCGGATG GAATTCATTC AACCGGCAAC	2760
GAGGTTTACA GATGGCTGGA TGAAAACgC TATCGCTTTG GTTTTATTCG CCGCTACCCG	2820
CCAGACAAGA CAGAGATAAC CGGTGTGAGC AACGAGCGT GGCATTACCG ATATGtCGgC	2880
ATCGAAGCTG CCACAAAGAT ATACCACCA GGGCTTTGCC TTGAGGAATA TTAAACACA	2940
GAAAAATGAG AAAAGGATAT AATGCTATGA ACAGAAAAAG ATTGACACAG CGCTTCCCGT	3000
TCCTGCTTCC AATGAGACAA GCGCAGAGAA AAATATGCTT TTATGCGGGA ATGAGATTTG	3060
ACGGCTGTTG CTATGCACAG ACGATAGGAG AAAAAACGCT TCCCTATTTG CTCTTTGAAA	3120
CGGATTGTGC GTTATACAAC CACAATACCG GATTTGACAT GATATACCAA GAAAACAAGG	3180
TGTTCAACTT AAAGCTGGCG GCAAAGACCT TAAACGGCCT ATTGATAAAA CCGGGGGAAA	3240
CCTTTTCTTT CTGGCGGCTG GTACGCCATG CGGACAAAGA TACCCCTAT AAAGACGGCC	3300
TTACGGTGGC CAATGGTAAG CTCACCACCA TGTCGGGCGG CGGTATGTGC CAGATGAGCA	3360
ATTACTATT TTGGGTGTTT CTGCATACGC CATTGACAAT TATCCAGCGC AgcGGTcACG	3420
TAGTAAAGGA GTTTCAGAG CCAAACAGTG ACGAGATCAA AGGGGTGGAT GCAACCATCT	3480

CAGAGGGCTG	GATTGATTTA	AAAGTGC	GAA	ACGATA	ACCGA	CTGCACCTAC	CAAATATGGG	3540
TGACCCTAGA	TGATGAGAAA	ATCATCGGTC	AGGTGTTTCGC	CGACAAACAG	CCTCAAGCAT			3600
TATACAAAAT	TGCAAACGGC	AGTATTCA	AGT	ATGTCCGTGA	AAGTGGCGGG	ATTTATGAAT		3660
ATGCCAAGGT	TGAACGGATG	CAAGTTGCCT	TAGGTACCGG	GGAAATAATA	GATTGCAAGC			3720
TGCTTTATAC	AAACAAATGC	AAAATCTGCT	ATCCCCTCCC	GGAAAGTGTG	GATATTCA	GG		3780
AGGCGAACCA	ATGAGAAAAA	GATGGGCAT	TACTGTTTTT	GGATGCGAGC	AGGATGAGGC			3840
AAATGCTTTC	CGCACCTTAT	CACCAGATTT	TCATATTATC	CCTACGCTGA	TCAGTGATGC			3900
GATATCGGCA	GACAAACGAA	AATTGGCCGC	TGGCAATCAA	TGCATTAGCG	TAGGCCATAA			3960
GTCCGAGGTT	TCCGAGGCGA	CAATTCTTGC	GCTGAGAAAG	GTCGGGGTAA	AATACATTTT			4020
TACCCGCAGC	ATCGGCTGCA	ATCACATTGA	TACGACTGCC	GCCGAGAGAA	TGGGGATCTC			4080
GGTTGGCACA	GTTGCGTATT	CGCCGGACAG	CGTTGCGGAT	TATGCTTTGA	TGCTGATGCT			4140
GATGGCCATA	CGGGGTGCAA	AGTCCACCAT	ACACGCCGTG	GCGCAACAAA	ATTTCA	GG		4200
GGATTGTGTC	CGGGGGAAAG	AGCTGCGGGA	TATGACTGTG	GGAGTTATTG	GAACCGGCCA			4260
TATAGGGCAA	GCGGTCGTCA	AAAGGCTGCG	GGGATTTGGA	TGCCGTGTGC	TAGCCTATGA			4320
TAACAGCCGA	AAAATTGAGG	CAGATTATGT	CCAGCTTGAT	GAGCTTCTAA	AAAACAGCGA			4380
TATTGTTACG	CTCCATGTGC	CGCTTTGTGC	GGATACCCGC	CATCTGATCG	GCCAGAGGCA			4440
AATCGGAGAG	ATGAAGCAAG	GCGCATTTTT	AATCAACACT	GGGCGCGGGG	CGCTTGTCGA			4500
TACCGGGTCG	CTGGTGGAGG	CAC	TGGGAAG	CGGAAAGCTG	GGCGGTGCGG	CACTGGATGT		4560
GTTGGAGGGC	GAGGATCAGT	TTGTTTATAC	CGACTGCTCG	CAGAAAGTGC	TTGACCATCC			4620
CTTTTGTGCG	CAGCTCCTAA	GGATGCCAAA	TGTGATCATC	ACACCCCAT	CGGCGTACTA			4680
CACCGAGCGT	GTGCTGCGAG	ATACCACAGA	AAAAACAATC	AGGAATTGTC	TTAACTTTGA			4740
AAGGAGTTTA	CAGCATGAAT	AAAATAAAAG	TCGCAATTAT	CTTCGGCGGT	TGCTCGGAGG			4800
AACATGATGT	GTCGGTAAAA	TCCGCAATAG	AAATTGCTGC	GAACATTAAT	ACTGAAAAAT			4860
TCGATCCGCA	CTACATCGGA	ATTACAAAAA	ACGGCGTATG	GAAGCTATGC	AAGAAGCCAT			4920
GTACGGAATG	GGAAGCCGAT	AGTCTCCCCG	CCATATTCTC	CCCGGATAGG	AAAACGCATG			4980
GTCTGCTTGT	CATGAAAGAA	AGAGAATACG	AAACTCGGCG	TATTGACGTG	GCTTTCCCGG			5040
TTTTCATGG	CAAATGCGGG	GAGGATGGTG	CGATACAGGG	TCTGTTTGAA	TTGTCTGGTA			5100
TCCCCTATGT	AGGCTGCGAT	ATTCAAAGCT	CCGCAGCTTG	CATGGACAAA	TACTGGCCT			5160
ACATTCTTAC	AAAAAATGCG	GGCATCGCCG	TCCCCGAATT	TCAAATGATT	GAAAAAGGTG			5220
ACAAACCGGA	GGCGAGGACG	CTTACCTACC	CTGTCTTTGT	GAAGCCGGCA	CGTCAGGTT			5280
CGTCCTTTGG	CGTAACCAAA	GTAAACAGTA	CGGAAGAACT	AAACGCTGCG	ATAGAAGCAG			5340
CAGGACAATA	TGATGGAAAA	ATCTTAATTG	AGCAAGCGAT	TTCGGGCTGT	GAGGTGGCT			5400
GCGCGGTCAT	GGGAAACGAG	GATGATTTGA	TTGTGCGCGA	AGTGGATCAA	ATCCGGTTGA			5460

GCCACGGTAT CTTCCGCATC CATCAGGAAA ACGAGCCGGA AAAAGGCTCA GAGAATGCGA	5520
TGATTATCGT TCCAGCAGAC ATTCCGGTCG AGGAACGAAA TCGGGTGCAA GAAACGGCAA	5580
AGAAAGTATA TCGGGTGCTT GGATGCAGAG GGCTTGCTCG TGTTGATCTT TTTTGCAGG	5640
AGGATGGCGG CATCGTTCTA AACGAGGTCA ATACCCTGCC CGGTTTTACA TCGTACAGCC	5700
GCTATCCACG CATGGCGGCT GCCGCAGGAA TCACGCTTCC CGCACTAATT GACAGCCTGA	5760
TTACATTGGC GATAGAGAGG TGACCCGTAT GGAAAATGGT TTTTGTGTTT TAGATGAAAT	5820
GTTGCATGGT GTTCGTTGGG ATGCCAAGTA CGCTACATGG GATAACTTCA CCGGAAAACC	5880
AGTGGATGGG TATGAGGTGA ATCGCATCAT CGGCACAAAG GCCGTGGCGC TTGCTCTGCG	5940
CGAAgCACAA ATCCATGCGG CAGCCCTTGG CTACGGCTTG CTTTTATGGG ATGGATATCG	6000
GCCAAAATCT GCGGTGGACT GTTTCCTGCG TTGGGCGGCG CAgcCGGAgG ACAACCTCAC	6060
AAAAGAAAAA TATTACCCCA ATATTGAGCG AGCCGAGTTG ATTACAAAGG GCTATGTGGC	6120
CTCACAATCC AGCCATAGCC GTGGAAGCAC AATTGATCTT ACGCTCTACC ACTTGATAC	6180
AGGGGAACCTT GTTTCATGG GAAGCAACTT CGATTTTATG GACGAACGGT CGCACCATAC	6240
AGCAAAAGGG ATAGGGAATG CAGAGGCACA AAATCGAAGA TGCTTGCGTA AAATCATGGA	6300
AAGCAGCGGA TTTCAGTCCT ATCGCTTTGA ATGGTGGCAC TATAAGTTGA TTGATGAGCC	6360
ATACCCCGAT ACCTATTTTA ATTTGCTGT TTCATAATGA AAGTATTGA TTTTCTAATT	6420
ATGTATAAGT TGGCTACAAA TTAAGTAGTA TTTCATCAGA CCAATTACTC TCTTGTTTAC	6480
AGAAAAATTC TGCGCTGATG GAATCTGCTT TATTATGCGG GCGAAAAATG AAATTGACCA	6540
TATTTTTTTCa GAACTTTACT CTGTACCGaA TTGCCTGCAA AAGCCTTATT TTAAGCTGAA	6600
AGTTCAGGAA TTGCTTTTGT TTTTGTGTAT GCCCCTCGTG ATTTGTACAC CTATCTTAAT	6660
TGGCTTTGCA ATTCTCATTC CGTATCTCTG CTTTAAGAAT TTGGaAAAac GAAGCATTGT	6720
GAATCGGCTG CGGGCAGAGC AAAAAGAGAA CCAGCAGAAA CAAGTCGTTT TTGCTCTGCT	6780
GATTCACCTG GAACTGTTT ATTCCGGTTT TCGTTGAAGG TCAAGTAGCT GCTCTGTCAG	6840
GAAGTCCAGT GTGTTCAAGCA GAATCTGCTG ATTGTCACGG TTGCATGACT GAAATTTTCC	6900
CATGAAACGC TGGAGTTCTT CATCCTCAAT AGAGTTTGAA GCTTGAAACA GGACATTCCG	6960
AGAAATTCCC AGTCGTTCAA TTAACCGGGA AAGTGTTTCA TAAGACGGAT TTTTCTTCC	7020
CTTTTCGATG TCCTGAACCG TTTTGACCGC CAAGCCAGAT TGATCCGcCA GTTCCTGTTG	7080
GGTCAAGTTA CATCGAACCC TTTCGGCTCG GATGTGCTA CCCAGAAATT TCAGGCTGTC	7140
TTCCGGCATG ATtGTTCAACC TCATATATAT TCTATCGTAC CAATTTGTAT TTTGGTACTC	7200
CCTTAGAGTT GTGTTCCAAT CAGTATTATC ATGCCGGTTT TTGAGTTTAA TATGCAATAA	7260
AGCCCTGTTG GAAGCAATAA AAAAAACAGC TTCTCACGGC aGGGCTAAT TGCCaCGCCC	7320
CCGTGGTTTG TCTGTAAAAT CATGGGGGTA TTTTATGTTC AAAAGAATAA TGGGAGGTAA	7380
TGCGTTAAAA AAAGTGTAGC TCCGCCGTGG GCATTATCCG CCCCCTAAG CGCACTATCG	7440



TTCCTATACA ATACATATCA TCATTCCTAC TACCTTATTC GGTCGCATTG CGCCAGGAAG	7500
CCTAAGAAGC TTCCTGGCCy TTTTGTtTGT CGAATTTTGT TGAATGATAA CGAGAATTGT	7560
CGATGTCGCA GCCCGCCTTC ATCGGTTCGC AGCCCATGAA CTGATGAAAG GAGTAAACGG	7620
CTATGACCAA AGAATATTGG GTAGCGGTCT TAGAACAAGA AGATCGTnTG ATAAGTAA	7678

## (2) INFORMATION FOR SEQ ID NO: 62:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4732 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

ATTTATTATA AGGGAGCGAn ATnCATGGC ACGCGTAGAA AGTtTGAAT TaGATCACAA	60
CACAGTAAAA GCACCATATG TTCGCCTTGC TGGCACAGAA CAAATGGTG ATGCGTTAGT	120
CGAAAAATAT GACTTACGTT TCTTACAACC AAACAAAGAT GCCCTACCAA CAGGCGCATT	180
ACACACGTTG GAACATTTAT TAGCAGTTAA CATGCGTGAT GAATTAAAAG GAATCATTGA	240
CATTCGCCA ATGGGTGCC GCACTGGTTT TTATATGATT ATGTGGGATC AACATTCACC	300
ACAAGAAATC CGTGATGCAT TAGTCAACGT TTTAAACAAA GTAATCAATA CAGAAGTTGT	360
TCCAGCAGTC TCTGCAAAAG AGTGCGGAAA CTACAAAGAT CATTCTTTAT TTGCAGCGAA	420
AGAATACGCA AAAATCGTCT TAGACCAAGG AATTAGTTTA GATCCATTG AACGTATTCT	480
GTAATCTTTA GCAATTAACG ACAGAGAAAC CCACTATTAA GCTGTGGGT TTTCTGTTCT	540
CTTCTTAAT TTAAACGGAG GTTTTTTATG AATCAAACAA TCGAACAATT ACTAAGTCAC	600
CGTTCGCTC GTCATTTTAA AAAGCAAGCG TTAACGGACA AGCAGAAACA ACAGTTAATT	660
ACTGCAGCTC AAGCAGGATC AAGETCCAAT TTTTACAAG CTTATACAAT TATAGAAATC	720
AAAGATCCTG AGCTACGCAG AGAATTGGGA CGTTTAGCGA ATTGTGAAGA CTATGTCGTT	780
AACACTGGTG TGTTCTATGT TTTTGTGCT GATTTGTATC GTCATGCAAC GATTTTGTCA	840
AAAGAGGGGC AATCATTGGA ACCCTTAAAG ACACCAGAAT CATTATTAGT AGCTGCCGTT	900
GATACAACGA TTGCTGCGCA AAACATGGCC ATTGCGGCTG AATCTATGGA TTTAGGGATT	960
TGTTATATTG GCGGCATTG AAATGACTTA GACACAGTAG CCAAACGCTT ATCTTGCCT	1020
GAATTAAC TG TACCACTTTT CGGATTAACA ATCGGTGTAC CAGAAACGTT AAATGGTGTA	1080
AAACCACGGA TGCCTTTTGA AAATATTTTA AGTGAAATC ACTACCAATC AGACAAATTG	1140
ACAGATATGC ACACATATGA TGAATTATTA AAAGATTATT ATGCCAGCCG CTCAAGTAAC	1200
GCCCAAACAG CGGATTGGTC ACAGAAATCT TTGTCTATT TTTCTTATAA TCGTCGCCCA	1260
GAAGTAAAAA TTTTCTTGCA AAAACAAGGA TTtGATGTCT AAACAAAAGA GCGTGGGCTT	1320
ACGTTATTTC AACGTAAGCC CACGsTsTTT TGTTAATTTT TATTTATCTT TACTAAATTT	1380
GTCTTTTAAG TCTTCAAAGC CTTCTTAAAC TTTATCAGTT AATTTTTCAG CGCCTTCTTT	1440

AAGGTCGTCG CCGGCATCAC GTGCTTTGTC TTTTACGTCA GCAAAGGTTG ATTGAGCTTT	1500
TCCTTCAAGT TCTTTGCCTT TATCATCCGT AACTTTGCCT TGCCTTCTT TTGCAGTACC	1560
TTCTACTTTA TCTTTTGCAT CGTCAAAACG TCCTTTTAAA TCTGCCATGC TAATCCCTCC	1620
TAAATAATTA ATTACATAAT AAGAGTATCA TTCTTAGCCA TTCCTTCAA ATATAATGCA	1680
TGAACTACGA ATAAATGGTT ATTTTGTAGC CATTAAAGCT TTTTGTGCCA AGGCAAAGCA	1740
ACCGATTGTT CCTGGATTAT CTTCTAAAAG AGGCGTCACC AAGTATTTTT CTAAAGGCGG	1800
TGTTTCCACA TATCCATTGA CTAATTCAGC AAATTTTTCG CGAAGTTTCG GCATCAAATG	1860
ACGTGTGTTT ATGACGCCGC CACCTAAAAT GATCACTTCT GGCGCTAATA ATAAAGTCGT	1920
ATTGTACGnC GtTGCCTAA ATAATAAGCT TCTAATCCC AAGTTTTATG ATCCTCTTCC	1980
AATAAATGTC CTTTTTTACC AGAACGACCT TCAACTGCTG GTCCTGCTGC AATCCCTTCT	2040
AAACAATCTC CATGATAAGG ACAATTTCTT GCATACGTAT CTTCAGGATG ACGACGAAC	2100
AACGCATGCC CCATTTCTGG GTGGCTAAAG CCTTCAATAA ACTCACCGTT TTGAATCGCT	2160
CCAGCGCCAA CACCAGTTCC AATTGTATAA TATACACAAC TAGATGTATG TTGCCCATTT	2220
CCAGCAACAT ACTCACCATA TGCCGCAGCA TTCACATCCG TTGTCCAAGC CATTGGCACA	2280
TCAAAATGTT GTTTCATAGT TCCTAACAAAG TCAAAGTTAC GCCAAGCTAA TTTTGGTGTA	2340
GAAGTGATAT AACCATACGT AGGAGAATCA ACGTGAATAT CAATCGGACC AAATGAACCA	2400
ATCCCAATCG CTTTTAAAGG ATATTGTTGG AAAAATTCTA TTACTTTTTT CATTGTTTCT	2460
TCTGGGGTTG TTGTGGGAAA ACTGACACGT TCTACGATGG TCAAATCATC TGTCCCAACG	2520
CCACATACAA ATTTTGTGCC ACCGGCTTCG ATACTTCCTA AAAGTTTTTC TGTCATTGCA	2580
ATATGCTCCT TCGTTTCCAT TTTACAAGTA AGCGGATACA TTACTTTTGA TTAAATAACT	2640
AACATACACG TTTAGTTTAC CACTTTTTTG TCGTTTCGAC AGAACAATTT AACCAATTCA	2700
GTCAGGTTGC CCAATGAAAT CTGGCAAGTT GCTAAAGAGC ATTTTCCTAA AACAAAAAAA	2760
CGAGCATTCA CGGTTATCTA ACAATGAATG CTCGTTTTTG TTTATTGATT CTTTAGCGAA	2820
AAATATAACA ATAGTCCGAT TGCAATCCAG CCACCACTAA GTAAAAGCAT TAACCAACTT	2880
AAAAGCGTTT CTTTTAGAAT ACTTCTGGGC TCTTTTGCAT GCTGAATAAA ACTCATACTA	2940
GTCACCAAAG AAAAGAGGAA AAACATAAAA CAAATGACCG CTAAACCGCC AGACAACCAA	3000
AACCAACTAG CCATTATTCT CACTCTCCTT TTCGTTGTT TTTTGATAAT TATTGCGCCA	3060
AATGACACTA TAAATTGCCC CAATTGTGAT TACTAACAAC GTGTAGAAAG TCGGTAACCA	3120
CCATGTGGGT AGGACCCATT TGACCGTAGC TTCTTGATTC AATTTTTTTG CTTACGAGT	3180
GTTAATCGTA AATTCTTGTT GAAAAGGCCA AGACTGTTGT TCATCTTGTT GTAAATAGC	3240
ACGTCCTGTA AATATATAAG TCCAGCATC TAGCGTTGT TTACCCAAAT CAAGATGAAA	3300
ATTCATGGAA CTATTTGGGG CAAAGCGAAC ATTTTTTAGT GTCCGTTGTG CAATTTTTTT	3360
AGAGGATCCT TTTTTTATCA CTTGACCTTC TAAACGAACT TCTGGAAAAA TAGCAGCATA	3420

TGGATTTTGA ATGACCGCCT TTAGCGATTT GGCTCCTAAA GCAATTGTAG GTGTTACTTT	3480
CGCTAATTTT AGTTGGTTCA CCTCTGTTAA AGGAACATTC GCAGCACTCC TCATGGCAAT	3540
CGCAACATTA AACCCGTAGC GACTGGTTAA TCCCGTTGTT TGGTCTCTT CCTCTGTTTT	3600
TTCATCTTTG AGATGCGCAT AAACCCCGCC CAAAACAACT CCTTTGATAG GCTGATCAAG	3660
TCCCTTCACC TTCATCTTTG CTGATACTGT TTGATGAGGC GGCACACTCA CTTCTCATT	3720
TTCTAAAGTA ACCAAGTCTG TCAGCGGAAC CTTTAACGAT GAATCAAGTG GTTTATCAAA	3780
ATCATTGTAC ACAATCAAAC CATTATCATT TGTTGaTGCA TTTGCTAGAG TCAATACGAT	3840
GGaCAATTTT TCGTCTGTCT TATTTTTTAT GTCTAAGACA ATAGTGAAG ATTGCTTTGG	3900
TGCCAGAATT AAATCATAAT AACTAGATGT TCCTGACACT TGATTTTCTG GTAAAACAGG	3960
CGTTACTGTA AATTTAGTCT TCATTCTTC TGTAGCTTCT ACTTTTGGAA CAATAATAAA	4020
AAATGTGAAA AACAGAAATA GACAGAwAAC ACCGACTTTT TTTCTACTCA TTTCTTCAAG	4080
ATCTCCTTTA TAAAGCTGAA TGTACCTACT AGTCTAACAA ATCAGCGAAA CAATGCTGAT	4140
CAATCTTTGA CTTTGGGAA GGATAATCTT GAAAAAGTC TCCTTGAAAA ATACCAAGGA	4200
GACTTTTTAT TaTTCACTAT AATATAATTC TTTTTTTATC TGTGTGGTTG ATACTTCTGG	4260
CGTTCGTGGC AAATAAGAA CCTTTCGGT TGTTTCAACT TCTAGAAAAT CAAAAGCCCC	4320
TTGCCAATCA TCGCCCATGA CTAAAGTATC AATTTTAAAT TCAGCAATAT CGGAACTTT	4380
TTGTTTCCAA CTAGTTTCAG GAATAACTAA ATCAACATAA CGAATTGCCT CTAATAATTG	4440
TTTACGTTTT TCATATGAAA AATAACTTTG TTTCTTTTTT TCTAAATTAA AAGCATCTGT	4500
TGAAAGTCCA ACAATCAGGT AGTCCCCTTG TTGTTTCGCT TTTTCAACA AATTAATATG	4560
ACCATAGTGT AGGAGGTCAA ATGTCCCGTA AGTAAGTATT TTTTTCATAA AAACCTCCAT	4620
TTGTACTTTG ATATTCTTC GTTAATTATc AATTTTtaTG AATAGAAAAA AGGCTACTCT	4680
TGaTTGaCTT TGaTGGTTTT TCGATGCAAA TTTTTATTT TnCGTAGCCG Tn	4732

## (2) INFORMATION FOR SEQ ID NO: 63:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7294 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

GGGaaAATwt TTTTAtgkGc tATGGcGTAt TACTAGTTAC gTTGTTGCCC TAGTGTCcT	60
TATTTAAAAt ATTTGGtGAA AAATCaGCGT TTATGGACcA TTGgATACGG CCcATTTCcA	120
GTGGGGAATG TTTTkgcCTG TGCTTTTTGG CTTTTTCATG CTGATTGGCT TTGATCAATT	180
TTTTATTGCG TTTcATGGAT TATTTTTTAA CAACGATGCT TGGTTATTTA ATCCGCTGAC	240
TGATCCGATT ATCAATGCCT TACCTGAAGC CTACTTTATG CATTGTTTTA TCTTGTTTTT	300

TATTTTATTA GAAGGCCT	TTTTGATAGG TATTTTAATC GGAAAACGTT CTTTAAAAAA	360
ATAAAGAGAG AAAAGTGAAC AATTCCAAAA	TTGCTCGCTT TyCTTTTTTA TTCAaTcAG	420
TTTTAAAAAC TAGATTGAAT AATTGTCAAA	AGTATGATAT TCTTTTTTAG AAATGGTTAG	480
GAGGGTTTTT ATGGACGAGA TAGAAAAATC	ACTACAAAAT TGGGGCCAAC AATTGGAAAA	540
CGTTCATTG CCAAGATGGC ATGAGTTACC	AGATATTGAA TTATATATGG ATCAAGTGAT	600
TACTCTGATA GAAAAATACT TATCCCCATT	GATTACACTG GAAAAACATA CTTTGTTAAC	660
GTCCTCAATG GTTAATAATT ATGTCAAACA	CGGGTTAATC CCGGCCCCAG TGAAAAAGCG	720
CTACAATCAA AAGCATTTAG CGTTTTTAAT	TGCGATTACG TTGTTGAAAC AAGTGTTAAC	780
ATTGCCAGAA ATTAAGCAAG GCATTTTATT	CCAAGGAGCT ACGGTTGGTA TTCGTGAAGC	840
gTACAATTTA TTTTGTCAAG AACAGGAACG	AGCGATTAT GTGATTGCTG CTCAAGCGCA	900
AGAAAAAGAA GTTCAGGCCA AATCACAAGA	ACCTATGGGG ATTGAATATT TGGCTGTAA	960
AGCAGCGACA ATGTCTTTTG CTACTAAAAT	GTTTAGTGAA AAAGTGATTG AATTGGAACA	1020
AGAATACTTA AAAGAAATGG ATGAGATGAC	ACATGAATAA AGAAAAATC GCACTACTAG	1080
TTGATTCTGG GACAGATGTT CCAGAGGCAT	TAGTCAAACA ATACGGCATG TATGTCTTGC	1140
CCTTACAGAT TATTTATCCT GAAAAAACCT	ATACAGATAA GGTAGATATT ACACCAGAAG	1200
AAGTGATCA ACGTTTAGAG AAAGAAATCC	CTAGTACTTC ATTACCTGaT GGTGCAACAA	1260
TCCAAGCTAT TTTTGACaAA ATTAAGAAG	CGGGTTACGA AAAAGTCTTA GCGGTGACCA	1320
TTTCTAGTGG TTTAAGTGGA ACGTATAATG	TAGTTCGCTT GCTTGGAGAA CAAACCGAAG	1380
GACTAGATGT ATTTGTTTTA GATACAAAGA	ACATTGGTAT CGGGGCTGGC ATTCAGGCCA	1440
TACGAGCAGC AGAATTAATT GAAACAGGAT	TAGGATGGCA AGAATTACAA CAAAAATTAA	1500
CAGAAGAAGT AGCCAACGCA AAAGTATTTT	TTAATGTAGC CACTTTAGAA TATCTACAA	1560
AAGGTGGACG AATTGGTTTA GTCATTCTA	TTTTAGGGAA TGCGCTAAAA TTAAACCCCA	1620
TTATTTCTTG TAATGAAGAA GGCATTTACT	ATACGGTGGC TAAGTCTCGC GGCCGTAAAA	1680
AAAGTCTTGA TAAAACGTTT GAGTTGGTGA	CGAACTTTAT AGGTGAGGCA CCCCCTTCC	1740
GTTTGGCAGT CGCTCACGGT GCGGCAGAGG	AAGAAGCmAA AGCmATGATG GAACGCTTGA	1800
AAGCAGCTTT TCCACAAGCA GAAGAAATTT	ACTTTGGAAC AATTAGTCCG GCATTAGTTG	1860
TTCATACAGG ACCAGGACTT TTAGGTGTTG	GCATACAGTT ATTAAATGtT AtTAAATGAA	1920
AAATCCCCTC AAATTTTTTT GaGGGGrTTT	TTCaTTTAAG ATTyCyCTTT TyCATAGrAA	1980
TATCaTtTAG CyGGaAAAAC aACGrAAATT	wTwTTGaAAA TyTCACGAAT GTGTTACAAT	2040
CCATCTACAA AGGGGGAtTT TTTGTGaAAA	AAGCAATAGC AGAATGTTA GGAACATTTA	2100
TTCTTGCTT TTTTGGGACA GGAACAGCCG	TTCTAGGAAA TGGTATGGAA GGGATTGGAA	2160
CAACGGGGAT TGCCTTAGCA TTTGGGTAA	CAATCGTTGC TGCTGCCTAT AGTATTGGCA	2220
CAATTTCAGG TGCACATCTA AACCAGCTG	TTTCTATAGG AATGTGGCTT AATAAACGAA	2280

TGACAACAAT	GGAAC TAATC	TACTATGTAG	TGGGACAAAT	TGTTGGTGGC	TTGATTGCCT	2340
CATTTGCTTT	ATTAAGCATT	CTAAAAGGTG	CTGGTAAAAAG	TATTGAGAAC	TTAGGTCAAA	2400
ATGGCGTAGG	GGAATTAAGT	GTTGCTGGTG	CATTGACAGT	CGAAATTATT	TTAACATTTA	2460
TTTTTGTTTT	AGTTGTCATG	ACAGTTACAA	GTGCTAAGAA	AGGCAATGCT	TCATTAGCAG	2520
GAATCGTAAT	TGGGTTAACA	CTAACAATGA	TTCACCTAGT	CGGCATTCCCT	TTGACAGGAA	2580
CTTCTGTCAA	CCCAGCTCGT	AGCATCGCGC	CAGCCGTTTT	TGCGGGTGGT	TCAGCGTTAT	2640
CTGAGTTGTG	GATTTTTATC	GTTGCGCCAT	TAATTGGTGG	ATTATTAGCC	GCTCTTGTAG	2700
CAAAATTCGT	CTTAGATACG	GAAGAATAGA	TTAGCGAAAG	AAGAATCTTT	GATAACAAGC	2760
GTCAATTGTT	GTCAGAGATT	CTTTTTTTGT	TATACTAGGA	GAACCTAAAG	GCGCATTTTT	2820
AACCAAAGAT	GAAAACATCA	TCAAATAAG	GTATAATAGG	AAAGTATCAA	ATAAGATTAG	2880
AAGGAAAAAG	AGGTGTCTCT	TTGTGAAGAA	AATTTTAGTA	GTTGATGACG	AGAAGCCAAT	2940
TTCAGAGATC	GTTAAATATA	ATTTGGTTAA	AGAAGGATAT	GAAGTATTTA	CTGCTTATGA	3000
TGGAGAAGAA	GCACTTGAAA	AAGTGAAGA	AGTGAACCA	GACTTAATTA	TTTTAGACTT	3060
AATGCTCCCT	AAAATGGATG	GCTTAGAAGT	CGCGCGAGAA	GTGCGCAAAA	CACATGATAT	3120
GCCAATCATT	ATGGTGACTG	CCAAAGATTC	TGAAATTGAT	AAGGTTTTAG	GATTGGAATT	3180
AGGAGCCGAT	GACTATGTAA	CGAAACCATT	TTCAAATCGT	GAATTAGTTG	CTCGTGTAAG	3240
AGCCAATTTA	CGGCGAGGTG	CAACCAATGC	GAAAGAAGCC	GAGGTGACAA	CACAATCTGA	3300
ATTAACGATT	GGTGATTTAA	CCATTCATCC	TGATGCATAC	ATGGTCTCAA	AACGGGGTGA	3360
AAAAATTGAA	TTAACCCACC	GTGAATTTGA	GTTACTTTAT	TACTTAGCAA	AACATATCGG	3420
ACAAGTGATG	ACTCGTGAAC	ATTTATTACA	AACCGTTTGG	GGTTATGATT	ATTTTGGGGA	3480
TGTGCGGACA	GTGGACGTAA	CCGTACGTCG	TTTAAGAGAA	AAAATTGAAG	ATAGTCCAAG	3540
TCATCCAACG	TATTTAGTTA	CTCGTCGTGG	GGTTGGTTAT	TATCTAAGAA	ATCCTGAACA	3600
GGAGTAATCA	GGTATGAAGA	AAAAAGTTCA	CTTTTTTCAA	TCTGTTAATT	TTAAGATTGC	3660
TTTATCATTT	ATTCTGCTGT	TACTGATTGC	GATTCAAATC	ATTGGTGGCT	ATTTTCATTCG	3720
TGAATTGGAA	GCCACTACAA	TTTCCGATTT	TAAAAAAAAT	ATGGATTCCC	AAGTTGTCCA	3780
ATTGTCAAAC	ACGTAAAGTA	CGCAGATGAG	CAACAAAGAT	CTCGAACGTA	GTGACGTTGA	3840
TGCAAATTTA	AAAAAAGCGT	TATCTGATTT	TTCAAATGCA	GATATTTCTG	AAGCGAGAAT	3900
TGTCGATGAT	AAAGGGATTA	TTCGGGCAAC	CAATGATTTA	AATCAACAAA	ATATTATTGG	3960
GAAAAAGAAT	GATTATCGTG	ATTTAAATGA	CTTTACGAGT	AAAAAATATC	AAGCTTTAGA	4020
TAATGATAAA	CGCGTGATG	TGAATGTCCA	GCCGATTCAA	TCGCCTACTG	GAGAAACAGT	4080
GATTGGCGTC	CTTTATGTGA	AAAGTAATTT	AGAAAAATAA	TACCAAGAAA	TTACCAACAC	4140
AGCAAGTATC	TTTTTCACTG	CTTCTATTAT	TGCCGCAGCA	ATCTCGATTA	TTGTGACTTT	4200
ACTGATTGCA	CGATCAATCA	CGAAGCCGAT	TGGTGAAATG	CGCGAGCAAG	CCATTCGAAT	4260

CGCTCGTGGT GATTACGCTG GAAAAGTAGA AGTCCATGGA AAAGATGAAT TAGGCCAATT	4320
AGCAGAAACA TTTAATCAAT TATCAGAACG GATTGAAGAA GCACAAGAAA CAATGGAAGC	4380
AGAAAGAAAT CGTTTAGATA GTGTCTTAAC GCATATGACA GATGGTGTCA TTGCGACGGA	4440
TCGCCGCGGA AAGGTGATTA CGATTAATGA AATGGCCCTT TCATTATTAA ATGTAAAAAA	4500
TGAAAATGTG ATTGGGACCT CGTTATTAGA GTTGTTAGAT ATTGAAGAAG ATTACACATT	4560
GCGGAAGCTG TTAGAAGAGC CAGATGAAC TCTGATTGAT CGCTCAACGT CTGATCGTGA	4620
AGAAGACCAA ATGATTATCC GGGTAGACTT TACAATGATT CGTCGGGAAT CAGGATTTAT	4680
TACTGGCTTA GTTTGCGTAC TTCATGACGT CACAGAACAG GAAAAAACG AACGGGAAAG	4740
ACGGGAATTT GTTTCCAATG TTTCTCATGA GTTGCGAACG CCTTTAACAA GTATGCGTAG	4800
TTATATAGAG GCTTTGASTG AAGGAGCTTG GGAAAACCTT GAGATTGCGC CGAATTTCTT	4860
AAAAGTCACG TTAGAAGAAA CCGACCGGAT GATTTCGTATG ATTAATGATT TGTAAATTT	4920
ATCTCGGATG GACTCTGGGg ATACACATCT TCAATTAGAG TATGTGAATT TTAACGAATT	4980
GATTAATTTT GTCTTGATC GCTTTGATAT GATGATTGAA AATGAGCAAA AAAATTACAA	5040
AATTTGCCGT GAATTTACTA AACGCGATT TATGGGTAGAA TTAGATACAG ACAAAGTAAT	5100
TCAGGTTTTA GACAACATTT TGAACAATGC GATTAAGTAT TCGCCAGATG GCGGCGTCAT	5160
TACCTGCCGA CTAGTTGAAA CACATAATAA TGTCGTCTTT AGTATCTCGG ACCAAGGTTT	5220
GGGCATCCCT AAAAAAGATC TCGGGAAAGT CTTGAGCGT TTTTATCGTG TGGATAAAGC	5280
ACGTGCGCGA GCACAAGGTG GGAAGTGGTTT AGGTTTAGCA ATTTCTAAAG AAGTAATTCG	5340
GGCCATAAC GGGAGTATTT GGGTGGAAG TACAGAAGGT GAAGGATCGA CTTTCTATAT	5400
TTCACTACCA TATGAACCTT ATGAAGAGGA TTGGTGGGAA TGATGAAATT ATCAGAAATG	5460
ATTACAAGAA TTGGCTTGAT TTTGATGGTC ATTTTAAGTA TATATTTTTC AGTCAATATC	5520
TGGCTGAATT CTGCCAAAAA AATACCAGAA ATGAAGTCGG GAAGCCAAGT CACAACAGCT	5580
GTCAATGAAA AAGCCATTGG CGATGTCTAT TTACCTTTGC AATTGATTCG AATAACTGAT	5640
GGAAAAGCGA TGCAAAGTAA TCGTGAAACA TTAATTAGTA ATGTTCAAAA TGATATTAAA	5700
ATGGCTACGT TTGGTAAATT GACACAAGTT GTGACAAAAA ATGCAGAGCA ACTTAAGCGC	5760
TACAACCAA TGGAACAAGG CATTGAACTT CTTTATCAAG GtCCCTTTTT AATCtCgGAC	5820
TATGCTTCGA TTTATAATCT ATCCATTAAT TTTACTAACT TtAATGAGTT GACGGACCAG	5880
TATTTTACGA AAATTCAATT GGgAtTTTAA CGAAAATAAG ATACGTTTTT TAGATTATGA	5940
TCAAGCCAAC GTCTATGAAG CGCCCATGAC TGTCAAATAAG GCGCGCTTAA TGGGAATTAT	6000
CAATAAGAG GGATTACAAT ATCAAGACGT TTCCGAAAAT ACGCTAACCA AACAAGGGCA	6060
ATGTTATTTA ACCAATGATA TGAAGTTGAA AAAGTACAGT TATATCTTAG CTTGCAACC	6120
AGTTACTCGT TTTAGGAATG CTTTTTCAA TGAAACGGAA GATATCCAAA CCAATGAAGA	6180
CAGTCAAGAC TTAACCTACA CGAGTAAAGA AGAACGATTG TTTGCAGAAG AAAAAGTGGG	6240

503

AAAAATTGAT TTTAAAGGGA CCTTGCCAGA AGAGAATAAA CGGGACTCAA TCTATAATCA	6300
AAGCTTTTCT TATGTAAAAC GTTTAGGAAC CAATATGGGG AATTTGCGTT ACTTTGATCG	6360
AACGAAAGAT AGTGTCAATT ATCGGACTTT TGTGGAAGGT TTCCCAGTGT TCAGTAATGA	6420
TTTAAAGGC CAAGTGGATA TTCGCATCAC GAACAACGAT GGTGCTGCAC CAAGCGTAAC	6480
CATTAACACA AGTGTGAATA CGATCCAAGT GCCGATTCCT TCAGAAGAAG AAGTGACGCT	6540
GGAAAGCACG GAAAAATTGA TTAAGCGTTT AGAAACGGCT GGTGCTAAAA AGGAAAAAAT	6600
TCAATCGGCT GTTATCgGTT ATACGTGGCA GACAATTGAA GAAGTCAAAC AAGTCGTTGA	6660
TTTAAGCCG GAATGGTATG TCCTTTATAA CAACAATTGG TACTGCAA CGGACTTGGT	6720
AAAACAATTA CCAAGCTTGG AGGTGGGCTA ATGGATTTTA AACGAATTGA GTGGATTTTC	6780
TTCTTAGCAT TTTTAGGATT AAATATCTTT TTATTTGGAA TTTATCGGGA AGGTCAACAA	6840
GAGGAGAGCA ACGTCTCTTC TTCCAGCCAA ACAGAATCGA TTGAAAGTCG GTTGAAAAA	6900
GATAACATCT CGTATAAAGG GACACTTTCT TCAGAACGAT TGGAAGGTTA TTATTTAAGT	6960
GGCGAACAAA CCAATTTTTC TGCTGCTTTA AAAATCCAAC GTGAAAAGAA TAAAAATTTT	7020
TTGAGAAATG GGCTGCAAAT TGCGGATAAT ACTTTAACGA GTGTACCTAG TAAAACTAT	7080
TTTATTGATC CTAAGAAAAT TGATAAAGAT TTAAGTACCT TTTTAAATGA AAAAAATGCT	7140
TTATTATTCG GAGACGAATA TCAATACTTA CCAGAATTTT CTCATTTAAA AGAGCCGACG	7200
GCAGAAATG TGGCTGCACA ATCGTATAAA GGAATTCCTT TTAGAGACGA CACGGCAAAA	7260
TTAAGTATTT TAGCAGATAC GTCAGGTGAA TTAT	7294

## (2) INFORMATION FOR SEQ ID NO: 64:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3422 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

nATTTTCATCT AACTTTTTGA CGCCnTAACA CCCTATTATC ACTTTTTACA TCTTATTTTCG	60
ATTTAAAAAT AACGATAAAA AGCCCGTTCC TAACAGAATT ACTGCTGAAA CATAATAAGG	120
GAAAAAGTGA TTGATATCAA ATAAAGCTCC TGCTGCCATT GGTCTTAAAA TATTACCAA	180
ACTTGTAAT GTCGAATTTA GTCCGTTGaT GGTTCCTTGT TGATCTCCAG CATGTTTGA	240
TAAATAAGTA GTTACTGCTG GTCTAAACAA GTCAAACGCT AAAAAGACAA TAAACGTTGa	300
AAATACAACA ACTAAATTAT TTTTGTAA GGCAATCACG GCAATAAAAA TGGCACTTGC	360
AAAAAAGGTT AATTGGATTA AACCCATTTT ACCTATTTTT TGGACGATTG CATCAAAGAA	420
AAACAACCTGA CAAATCAACG CTAAAATACC ACTAACCGTA ATCACGATTG CTATTTCACT	480
TGTGGTAAAG CAAAATTAA TAGTCGCCAT AATACTATAA ATAGATTCGA ACGCTTGCAG	540
GCCAAAAGAG GAAATTAAGA TAATCACAAA TAATGAGGTA AACATTGGAT TTCTTAAAT	600

ATCCATAAAT GAACCTTTTT TCGCTTCAAC AGCGGCTAAA ATTCGTTTCT CTGGCTCCTT	660
CAAAACAGTT AATGTCAAAA TAAACCAAT AAACGCTAAA AATGCGGCTG CAAAAACGG	720
GGCGCGAATA CCTAAATAAG CaATAAAACC ACCAACTCCT GGTCCGATAA TAAACCACC	780
ACTAATTGCG GCTGACACAA GCCCCATCGC TTTCGGACGT TCAGAAATCG TGGTCATATC	840
TGCCACAAAG GCTGTCACTG ACGGCATTAA TAAAGCGGCG GCAATCCCGC CTAAACCACG	900
AGAAATATAA AAACCGCTTT TCGCTTGGGC TAAACCAAT AATAACTCAG AAATTGAAAA	960
CACCAACATG CCCGTTGCAA TCATTTTTTT CCGACCAATT TTATCCGAAA GGACACCTGC	1020
GATTGGCGAT GTGATTAATT GGGCAATCGC AAAAATAGAA ATCATCATTC CCATTGTGGT	1080
ACCAGAAAAA TGCATTTCTT CTTTTAATTG GGAATTACG GGAATGACTA GGCCTACGCC	1140
TAGAAAAACA AGAAATAAGT TAGAAATTGC TAAGTACATC ATACTATTTT TTTTCGTCAT	1200
AATTTGCTCT CTCCATAAAA AAACAGCCAT CAAAAATGGC TATTTCTGTT TCTTTGTATG	1260
TACTTTCTTT GACAGTGTTT CGGCTGAAAC ACTCGTTGAG TCATTGAGAT ACATTTTACA	1320
CTAATTCACT GAAAATTTCA CTAGTTTTTA CTTATTTTTT TAAAGGAAAT TGTTATTTTT	1380
TAAAGACATA ACTTAAGTAT CCAAGTATTT AATGAATCGC TTTTGTTTTT TATCAGAAAC	1440
ATAGCCAATG TGTTCATAAA ATTGATGTGC TGCTGTACGC TGTTCCCCTG AATTTAAGCG	1500
AATCGCAGAC AAGTTACACT TTTTA <sub>g</sub> CT <sub>c</sub> C GTT <sub>c</sub> TAGTTT CGTCATTAAt TGtTTACCAA	1560
TTCCACGGCC TTCTGCTTCT TTTGCCACTG CTAATCCC <sub>a</sub> G AACGTtCAAT AACGTTTGGG	1620
AATAAATTGT TTCATAC <sub>a</sub> CT TGCGCATGCA CATAGCCTAA AATCTTGCCt GTCTTCGTTT	1680
CTTCATAAAC ATATAAAAAC TCTCTGTCAC TGGCTAGTAG TAATTTCTCA ATCTGTTTAC	1740
AGGTACCGTC AACGGGAAct TCATAGCCCA ATTGTTCTTT ATTAATGATA GTTAACGCCG	1800
CTGCATCTGT TGGCTTCATT TTTCTAATCA AGGGTAACAC CCTCCTTTAC ATTCTTACTT	1860
AAATAATAAC GAAAGCATAA TATCCGCAA ATCATTTAAT TCTGAATCAT TGTATTGCTC	1920
TAACAAAACT GTTTCGCCTC GATACTCTTT TAAAGCTAGT TGTATCAGCG AATGATGATT	1980
AGTAGGAAAA TGTACTAGGC CCCATTCCCC GCCAGCCCTT TTTGATAAAA TTTTATTCTC	2040
TTGCTTAAAT GCCATTACTC TACACATATT TAAAATAACG TACATAGGTT GTTGAACAA	2100
TTCTTTTTTT GCCTCTTTCA CATCAGCTAC AATACTTTCA AAATATTTCT TTTGATCTAT	2160
CCAGCCAAAA ACGGAATCAA TCGGTTTTCC TACGAGTACT TTTCCATATA ATTTTGTAAT	2220
CATAAGATGT GCCACTAGAT CCTCATCGTA TCCCATCATA TCTTCAATAT ATTTATTTTT	2280
ATTAGTCAGA TATCTGGAAC GATGTATCGG AGAAAAATGT AAATCGAAGG GAATCGGATA	2340
ACTCAAATTT ATTGTATCTT tTAATTGtAG CACAkGGAAT TCTAtCCCTT TTTTGGkGc	2400
CCACTTATCT AAAATATTAA TTGTATAATC TAATAATTGT TTCTTCTCAT CAATAGTAAG	2460
CTTTTCATAA ACCACAATCA GTACATCCAA ATCACTTACT GAATGACAAT AGGTTCCAAA	2520
AGCATAAGAA CCATGAAGAT ATATTCCAAC TAAATTTTCT TTTATTATTT TCATATATCC	2580



ATTTTCCATT TGAGCTAATA ATTGATGAAC CACGATTATT CCTCCATTAT TGATTGACAA	2640
TTGCTATAGT TTTTTCGATA AATAAAGAAC TAACTCATCA TCATTAACAC ACCGTTTCGTT	2700
TGGTTTCAGC TGCTTATTTT GAAACCAGAC TCCTGAACCA TCTGGCACAT ATCCTTGTTT	2760
GATATATAAT CTTTGCCTG CACCATAACC AGGATGCAGG CCTACTCCCA AAGAAACAGT	2820
ATCTGCATAA CTTTGGCAA CATTCTCTGC ACGATTTAAT AATAACCGCC CCACTCCTTG	2880
TTTTTGAAAT GGTAAGAATA CATTGAAATC AGCTATTTCA GGATATAAAT TTTTAAAAGG	2940
TCCTTCCTTA GCTAATGGCA GTAACGTCAC ATAACCTAAC AGCGTTGCCT TTTCTCCGC	3000
TACAAAACC GTTCGTTGCT TGAATAATTG TTCTTCTAAA TAGTTCATTA AAATCTCTTG	3060
GCGACTTGGC CAACCTTGTT GGrcAAATTG AGTATCCAGC GCTAAGATAT CTTTCTCTTG	3120
CATTCTCTTA ATAACCACTA GCTCAGACAT CAGCATTTCC TCCCGAAAAG CGTTTTTCATT	3180
TTATAAACG AAGCGAATAA TTCTTCGCTC CTTTATTCAT TTTCAAACC ATCGAGTTGA	3240
TAATATTTGG GCAAAACAGC AACACATGTT TCTTCTGAAT TAACAGCAAT AACGAACCTT	3300
TGCCATAAGG nCCACGTCAC CGCTTCACCT AATTTCTCTT TCACTCGTTG nATCATTTCA	3360
GCTTGCGTTA AGTCTCGTGC AGGrAAATGT TGAACTTTAT CATTACTGAT TAAAAAATA	3420
AA	3422

## (2) INFORMATION FOR SEQ ID NO: 65:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9172 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

CCGGACCATC AAAATCACTT TTTTCCTGCT CTTTnAAGA AACGAAnGGA CCCATGTCTT	60
TTCAGAAAAG TTTGTTACAA TnACAAGCGT AGAGAGGATG TTAGGGTATG GATAAAATGA	120
AAGCTATTAC TTTTTTTGAT TTnGATGGAA CTTTGTTAGA TGGAACATCG CAAATAACTC	180
CTGAGATTAC CGCAGCAGTG GCGGCCTTGA AAGACAATCA AATTTTACCT TTAATTGCGA	240
CAGGACGGAC CTTGTGTGAG ATTCAGCCAA TTATGAAAGC CAGTGGCATT GATTCTGCGA	300
TTGTAATGAA CGGGCAATTT ATTCATTATG AAGGAAAAC GATTTATTCA GATGAATTTA	360
CAACAGAGGA ATGTGTTTCA CTGCATGAAC ATGTCAAACA GCGTGGTCAT GAATTGGCTT	420
TTTACAATGA GCGCCGTATT TTTGTACAG GGCATACTGG CACAGTGAAA CAAGCCTATG	480
ATTATATTCA TTCAGCTGTT CCAGAAATTG ATCCCACAGG TTATGAGAAT GATGCCGTCA	540
ACATGATGTT GGTCTTAAGT CAACATGGCG ATGATGATGA ATATTACTAT GAGCGGTTTC	600
CTGAATTGAC CTTTATCGC AATGGTCCTT TTTCAATAGA TATTGTTTCGT AAAGGGGTCT	660
CTAAAGGCTC TGGTGTCAA AATCTTTTAA ACACCTTGGG ATTAAACGGC ATTCCGACTT	720

506

ATGCCTTTGG TGATGGCATC AATGATTTGG CGCTGTTTGA GGCTGTGAT TATGGAATTG	780
CCATGGGTAA TGCGCGAGAA GAACTAAAAG AAAAAGCCAC CTTTATTTCA ACTAAAAATA	840
CTGAGAATGG CATTGTTAAT GGCTTGAAAA AATTTGATT ACTATAAGAG AAAACTACCT	900
TCTTCTATTG ATGTACTTTA GAAGAAGGTA GTTCTTTTGT ATGTTCAATT TTCTTTTCT	960
GCTGTGCACA GAGAGCTACT CCCACTACTT GAATCAGCAA AATCGCTCCC AACAAACAAC	1020
TTAATACGAT GTTTCCTGAT AGGATTGTtC TTCTGATTAA AATAATTAAT ACAACACTGC	1080
TAATTCGTA GAGTAACCAT TGGTTGTCCA TTTTATTGA CCCCTTTGTT TTTCTATGTT	1140
TAGGATAGGA TGTTtCCTAC TAAATGCAA GGnAAGATCC TAGATTCTTA CAACTAACCC	1200
ATAATTTTTA CTAATTCCTT ATTTGTTTC ATTTATTCTT AAGATTCATT TCATTTTtGg	1260
TAAAAAATCA CACTTTTTTT CTATCTCTGT TTTAAATTGC GTACAGTTCT TAAAAAAAk	1320
GATACACTAA AaGTGaACTC GaAAkGAaRG GrGCCATtCC TTTGGtTAAT AGCGAAAAAA	1380
GAAAACAATT AATTATTGAT ACATGCATTT TAGCTGGAAA AATTATGCTG GAAAGCGGAT	1440
CTGAAGTATA TAGAACCGAA GATACAATTA CTCGGATTGC AGCAAATGCT GGCGAGCCAG	1500
AGAGCGTTTG TTACACGACG GCTACAGGGA TTTTCGTGGG CTTTCGTtCC AGTAATTATA	1560
CGCAATTAGA AAACATCCCC CAGCGCTCGA TTAATTTGGA AAAAGTTAGC CTTGTCAATC	1620
AGTTATCAAG AGAATTTGCG CAAAAAGAGA TTACTTTACC AGAGCTTTAT CAGCGCTTAA	1680
CTCTATTAGA AACAGATACC CCCACTTTTT CAATTTCTGT ACGCTTGTTA GCAGCCGGGA	1740
TTGTTAGTTG TACGTTAATG TACATTTTtG GTGGCACTTG GcAAGATTTT ATTGCAACTT	1800
TTTTTGTTGG CGTGATTGGG TATGCCAGTT ATCTTTTTAC CCAAAGTTA TTCCAAGTTC	1860
CTTACCTCGA TTCATTGCA GCCGCTTTTG TCATTGGTTT GTTGGCCTAT TTGGCTGTGC	1920
ATTTTCACTT AGCGGTCAAC ATCGATAATA TCATTATTGG TGCGGTAATG CCGTTAGTTC	1980
CAGGTGTGGC GATTACGAAC TCTTTTCGCG ACATTTTAGC GGGCCACCTC ATCAGCGGGA	2040
CGGCGCGAGG GACAGAAGCT ATTTTATTG CTGGGTCTGT TGGCCTGGGA ATTGCCTTGA	2100
TTTTTAAATT ATTTATGTAA GGAGTGTCA AATGGAAC TAATTTTCATC TTCTCTTTAG	2160
TTTTTTAGCG ACTGTGACGT TTGGTATTAT TACGAACATT CCTCGTAAAG CACTGGTGGC	2220
TTGTGGCATT ACTGGCATGA CAGGCTGGAT GATTTATTAT GTCTTAACGC AAACTTTTGA	2280
TGCTAGTCAA ACCTTTGCAA ACTTTTtTAGG AACGGTCGGG TTAGGAATCG CTAGTATCTT	2340
TTTTTCCCGT TACAAAAAAA TGCCGATGAT TATTTTTTAT ATTCCGAGTT TGGTTCCTTT	2400
AGTGCCAGGA GGCCCTGCCT ATCAAGCAGT TCGGAGCATT TTGTTAGGAA ATATTGATGA	2460
CGGTTTGCAG TTGATTTTAA AAGTTGTCTT TACTGTGCA GCCATCGCAG CGGGCTTTAT	2520
GGTGACCTCG CTACTGGAAC GAATTGTCAA ACGTTTTTTC CCCAAGCGAC TCTTCCGATT	2580
GGTCAAAAAA TAAAAAACA AGCTCATTTT CTGAGCTTGT TTTTTTATTC TACTTCTTCA	2640
AAGAACAGCA ACAAATCTTG TAGCGTTAAT TGTGCTTTTT CAGCAGCGTT TAAATCTTGA	2700

ACAATTTTAC	CTTTTGTAA	AACAATTAAA	CGATTGCCAT	ATCGTAAAGC	ATCTTCCATT	2760
CGATGGGTAA	TCATTAAACA	AGTTAAGTTC	CCTTCTTCAA	TCCGTTGTTC	CGTTAAGTGC	2820
ATCAACTGCT	TGGATGTTTT	GGGATCTAAG	GCAGCAGTAT	GCTCATCTAA	AAGTAATAAT	2880
TCTGGTTTTG	TGATAGTGGC	CATTAATAAA	CTTAAGGCTT	GGCGTTGACC	GCCAGAAAGG	2940
TTACCTGTCT	GTGTATCCAG	ATGATTTTCC	AAGCCATTGC	CAACTTCTTG	GCAGATTTTT	3000
GTAAACAAGG	CGCGTTGCTC	ATTTATCTTC	CGTAATGTTA	AGGGACGTTT	TTGTCCGCGT	3060
TTCATGGCAA	GCAATAAATT	TTCTGCCACA	GTCATTCTTG	GTGCTGTCCC	CATTTTAGGA	3120
TCTTGAAAAA	CACGCGCTAG	AGATTTGGCC	CGTTTGACTT	CAGAAAGATT	GGTGCGATTG	3180
ACTCCTTCAA	TTAAAAATTG	TCCTTCTGTT	AGTGTCAAAG	TTCTTGCTAA	ACTATTGAAT	3240
AGCGTACTCT	TGCCTGCGCC	ATTTCCACCT	AAAACAGTAA	TGAAGTCCCC	AGGATAAATG	3300
GCTAAGTTCA	CATGATCCAA	AATCACTTTT	TTTTCAATCA	TCCCATTATC	AATGATTTTT	3360
GTGCGATTTT	TTAATTCAAC	AACTGGTTTC	ATTTGTTTC	CCCCTTATCA	GAAAAGCCGT	3420
GTAGTCGTAA	TGCTTTTTTC	AATTGTGGGA	TCATTAAACA	GATTGCTAAA	ATCACTGCTG	3480
AGAAGATTTT	CAAATACGTC	GTATCAAAGC	CTAATTTAAT	CACTAATAAA	ATCAATAGTT	3540
GATAAATAAT	ACTCCCTACA	ACAATCGCCA	TCAGCCGTTT	ACCAAACGTT	AATTCTCCAA	3600
ACAGTACTTC	ACCAATAAAt	AGTGAAGCCA	AGCCAATCAC	AATGACACCG	GTTCTTTTAC	3660
TTACATCCGC	ATATCCATCA	tTTTGGGCAA	TCAAAGCGCC	aGATaAGCCA	TCAAGCCATT	3720
CGaTaAAACC	AaTCCTAAAA	TTTTCATGCG	ATCCGTTTGA	ATGCCAATCG	AGCGAGCCAT	3780
TTCTTCATTG	TCACCTGTCT	CAATATAGGC	TTGTCTTAAA	CTAGTGTTAA	AGAAAAATAA	3840
CAGTAAATC	ATTACCACAG	CCAAAACACT	AATCCCTAAA	AAAATAGTAT	CATAATAATC	3900
GGGTAAGGCC	ATTTTTTGAA	AAACATCTTG	AATCTTCGGT	TGGTTTAATA	AAGACTTATT	3960
TGGCGACTTC	ATCACAAATA	AAATCACGGA	ATTTAAACCA	GACATCACTA	AAATTCCTGC	4020
TAAAATAATT	GGAATTTTTC	CTTTTGTAAA	CAGTAAACcA	GTCACTAAAC	CGGCAAGCAT	4080
TCCTGTCCG	ACCCCTAAAA	GAGTCGTAT	AAATGGTGAA	ACCCCTGTTG	TAATGGCAGT	4140
TACACAAACC	GCCCCCCTA	ACGGAAATGA	ACCTTCTGTG	GTCATATCTG	GAAAATTTAA	4200
AATACGATAA	GTCATAAAAA	TACCTAGTCC	TAACAAGGCC	CATAACATTC	CTGTCCAAT	4260
TGCTGAAACA	ATCATCGCTT	CTCCTACTTT	CTTTGATAAT	TAGTTAATCA	ATTTCGATTT	4320
TTTTGCTAAC	TCTGCGGGAA	TCGTAATTCC	TAACTTTTGA	GCTTGTTTTT	CATTGATAAT	4380
TGTGTCGCCT	GTATTAAAGG	TATAAATGGG	TGTAGTTGCT	GGTTGACTTT	TTCCAGACAG	4440
AACCTCTGCT	GCCATTTTGC	CGGCTTGTA	CCCCAACGCA	TGTTGATTAA	TCCCTACAGT	4500
TGCTAAACCG	CCTTGTTCTA	CCATTGTATC	CACCGAAGGA	ATAATCGGCG	TCTTGTTTCG	4560
ATTTGCTTCT	CCGACTACCG	TTTGCATGGC	ATTAGCAATC	GTATTATCAG	TTGGTATATA	4620
AATCACATCT	GTTTCCTTAG	TCATCACTTG	AACCGTTTGT	GCAATTTTCAT	TACTTGATGG	4680

CACAGCATAT	GTTTTCACTG	TCAATCCTTT	GGCTTCAGCC	GCTTTCTGCG	CTTCAGCTAC	4740
TTGGTATCGA	GAATTTTCTT	CTGAGGAAGC	ATAGAGAATA	CCTATTTTTT	TGGCAGCTGG	4800
TATCAATTCT	TTTAGTAAAT	CCATCTGGGC	TGACAAAGGA	GATTGATCAC	TAACACCCGT	4860
GATATTACCG	CCTGGCTTTT	GATTGTCTTT	CACCAATCCT	GCACTAACAG	GGTCCGTAAT	4920
GGCGCCTAAA	ATGATTGGGA	TTTCTTGGGT	AGTGTTGGCT	AACGCTTGCG	CAC TGCGCTG	4980
GCGATGCCAA	TTAAAACGTC	TGCCTTTTTT	TGTACTAATT	GTTGACTCAT	GATTGCTAAC	5040
TTACTCTGGT	CGGCTTGACC	ATTTTGAAAA	ATAATTTCTA	AGTTTTTGCC	TTTGT CATAG	5100
CCATTTTCCT	TTAATCCTTC	TTCGACACCT	TTATAAATTT	CGTCCAATGC	GGGGTGACTG	5160
ACAAACTGCA	AAACCCCTAC	AGTAACTGTT	TCTTTCTTTG	AATTCGTCAC	TTTTTGATTTC	5220
ACAAAAAAG	CACCGATTAA	ATAAATAGCT	AATAAAGACA	CGACAAACGT	TAAACGTTTA	5280
TTTTTCATTT	TGCATTCTCC	TAAGTCCCT	TTGTTTGGC	AACTTTTTCT	TTTAAATAAT	5340
AAAACCATTA	TCTATAGGGC	GTTTCTTCCG	TTGAAAGTGA	AACTGAGATT	CGTAATCCTC	5400
TGCTTAGTCA	AAATCCAAGG	AACACTTTTT	TACAAATGTC	AGGATTTAGC	TGCGACGAGG	5460
ATTGCAACTT	TGGTGAGAAT	CGCAGGAGTA	CTTTTTCTAC	GAATGGGACA	ACGTTCTCTT	5520
CGTTCACCTT	GTACTACTCA	TCACTCTTGA	AGGCCAAACT	GCGGCGAGTA	TACGCAGCCT	5580
GCATCTACTC	TGTCnCTTCG	TGCCAATCGT	AGTGCTTGTC	AGGATTTAAC	TGCGACGAGG	5640
ATCGCAACTT	TGGTGAGAAT	CGCAGGAGCA	CTTTTCTACG	AATGGGACAA	CGTTCTCTTC	5700
GCTCACCTTG	TACTACTCAT	CATCTACTCT	GTCACTTCGT	GCCAATCGTA	GTGATTCGTA	5760
GCAAAAAAAC	AACGACTTAT	CCTAAGTGT	TAGGAGTGGT	CGTTGTTTTT	ATGCAAAACA	5820
AAAACCACAT	AGATGGGTCC	TCTATGTGGT	CGTTTTTTCT	TCTTGAATAT	AACCAGCATA	5880
GATACACTTT	TTAGAATAAG	TCTTCTAAAG	TTGTATCTAT	GAAAGAATCA	TACATACAAC	5940
TATCTACGCC	AGTTTTGCCA	TGATTGGTGC	ATTTCTTTTA	CGTTAGATAA	TTGTGTCATG	6000
TATTTTCCCT	CCTAGTTATG	TTGTTAAAGG	TAGTCTACAC	TATATTGATT	TTGTTCGTCA	6060
AGAAAAAGTT	TTTATTTTCT	TGCATTTCTT	TTACTTCTAA	AAAACGAGCC	ACTCTTATGA	6120
GAGGCTCGTT	TTTCTTTATT	TCAGCATATT	TACCACTTAA	TTGTTTCGGT	TTCTTTCAAT	6180
TTTCACGGGA	ACTTTAACTG	CTTGTTTTTG	CGATTGAATC	AGCCATAAAC	TTACTAAAAT	6240
CGTCATTGAA	ATTACTAATG	CCATCGAAAA	TCCCCTCTTT	CTAATTCCTT	AAAAC TTTCT	6300
ACAATCTAAG	TTTAACATAG	CTAGCTATTA	TTCACCAGTT	TCTGGTACTG	TATCAGTGCT	6360
TTCTGTCTGC	GTTTCAGTAG	ATTGAGCAAC	GCCAGTCGTG	CTGGTCGTGC	TCTGTGTTTC	6420
CGCGGTCTTT	GCTGCTTCTT	TTTCAGCTTC	TTGCGCCGCT	TTTTGCTGGT	CTTTC AAAAC	6480
GCCAGCCATT	GGTAATCGAT	TTAATTCTGG	TATTAATTGT	GTTTTTAAAT	CCATAGATTT	6540
TTGACAACCC	GTCCGTGTCT	TTTCTTTGCC	GTTTTTCAAAG	AAAACAAGCG	TAATCTCATC	6600
GCCGCTAACA	TTGTATTGCT	CTTGTGCTTT	GCCAGCATTA	TAAATTAACG	GATAGATATA	6660

CAGCGAACGA	TGAAATTCAT	TCATTTTATC	TGAATTAAAC	AGCTTCACCA	TTTGATCGTA	6720
TTGCTTGCCA	GATGGTTTAG	CAAACATGAC	TGACATGGCT	CCTGTTTCTT	GGATTTTCTT	6780
GTCTGCTGTT	TTGTAATCTA	AGACACTGAC	TTTCTCTGAT	TTGATGGGTC	CTGCGACTAA	6840
CGATTCAACT	TGAGCAGCTC	GAATATTTTT	TTGCATCGTT	GGAATAAAAA	TGATTGAGCC	6900
AATTATTAAT	ACTAATGCTA	CAATAGCGGT	AATCATTAAC	AGTTTTCTTT	GTCTTAATTC	6960
ATAAAGTGCC	TCTTTTAGTT	CTTCCATAAA	ATTATTCAAC	TTCTAATCGC	TCCTCGCTTT	7020
TAGTTCTTAT	CTATTATACT	AATCATTTTC	TTTCTTGCCCT	ACCCTTGCGC	TTTAATTACT	7080
AAAAAAACCA	GCAAACTTCC	CCAGCTTTCT	AAGGAAAACC	CAACTCCTTT	TTAACAAGCT	7140
TCTTTCTTAA	ATTCAAAAAC	TCCGCTATAA	TAAAGCTATT	AGTCGGAAAG	AAACGAGGGA	7200
TTTATTATGT	CATTGATGG	CGTATTTACT	CATAGCATGG	tTCATGAGTT	GaCAGAAACA	7260
CTCGTAAGTG	GCCGGATTTC	AAAAATCCAT	CAACCTTATG	AAAACGAAGt	CGTTTTAGTC	7320
ATTCTGTGCGA	AAGGTAAAAA	CCATAAATTA	TTATTATCAG	CACATCCTAG	TTATGCGCGC	7380
ATACAACTAT	CTACAATTAC	CTATAGCAAT	CCGGAACAC	CTCCTAATTT	TGTGATGATG	7440
TTACGAAAAT	TTCTAGACGG	TGCCATTTTA	GAAACGATTC	AACAAATAGA	AAATGATCGC	7500
GTGATTCACT	TCACCTTTTC	AAAAAGAGAT	GAAC TAGTG	ACTTACAAA	TATTGTTTTG	7560
ATTGTTGAAT	TAATGGGACG	TCACAGTACT	ATCGTTTTAG	TTAACCAATC	TTCTGGAAAA	7620
ATCTTAGATG	CGATAAAACA	CATTGGAATG	AGCCAAAATA	GTTATCGTTC	GTTGCTTCCT	7680
GGAGCAACTT	ATATTGAACC	GCCAAAACCA	ATGGGACTTA	ATCCGCTCAC	GGCCTCAAAA	7740
GAAGAAGTCT	TTGCGCTTCT	TTCTACAGCT	CCTGAATTAA	ACGGACGCTA	TTTACAACAA	7800
CATTTCCAAG	GTCTAGGCAA	AGATACAGCA	GACGAATTAA	GTGCCCCTTT	ACTTGACAAA	7860
CCAAATGAAA	AAATGGCCAT	CTGGACGGAA	TTTTGGTCCT	CAGTCACTGA	AGCTGTTGTA	7920
CCGACATTAA	CAGTTACTGA	AAAAAAAGAA	TATTTTACGC	CCATTCCTTA	TCAATCATTA	7980
GTTGGTGAAC	AAACAACGTA	TGCGACACTC	AGTGAATTAC	TAGATGCCTT	TTATTCTGGT	8040
AAAGCAGAGA	AAGACCGTGT	CAAACAACAA	GGCGGCGAAC	TAATTCGCAA	AATTGAAAAT	8100
GAACTAAAAA	GAAATCAAAA	GAACTTAAA	AAATTACAGC	AAACCTTAGC	CGATACAGAA	8160
AACGCTGAAA	ATTACCGTCG	CGACGGAGAA	CTATTAACAA	CCTTCATGGC	ACAAGTCCCA	8220
AAAGGAGCCA	CCGAGGTGGA	ATTACCAAAT	TATTATGAAG	AAAATGCACC	GTTGCGTATC	8280
TCGCTAAATC	CTGCACTTTC	ACCTAATCAA	AATGCACAAA	AATATTTTCA	AAAATACCAA	8340
AAATTGAAAA	ATGCTGTCCG	AGTTGTAAAA	ACACAAATTC	AGCAAACACA	ACAAGAAATT	8400
TCTTATTTAG	AGTCAGTTGT	GGCACAACCT	GAAATTGCGA	CACCAATGGA	CATTGAAGTG	8460
ATTCTGTGAAG	AACTGATTGA	ACAAGGCTAT	TTAAAACGTA	AAAAAAATAA	AAAACAAAAA	8520
CAGCCGAAAA	AAAGCCAACC	CGATCTTTTT	TATGCCACAG	ATGGCACACC	TATTCTCGTT	8580
GGAAAAACA	ATTGCAAAA	TGATCAATTG	ACTTTACGAA	CCGCTAAAAA	AACTGATTAT	8640

510

TGGCTCCATG CCAAAGATAT TCCTGGATCT CATGTAATTA TTEGTGACGC ACATCCTAGC	8700
GAAGAAACCC TTAAGTGAAGC CGCACTTCTA GCAGCTTATT TTTCAAATA TCGACTTTCT	8760
TCACAAGTGC CTGTTGACTA TGTCCAAGTA AAACATGTGC ATAAACCCAA TGGAGCAAAA	8820
CCAGGTTATG TGATTTATGA AAACCAACGA ACGTTATATG TGACGCCAAC AGAAGAATCG	8880
ATCAAAAAAA TCCAGcAAAA TAAGGCTTCT GCTTCCTAAT TGTATTAAAC TAAAAACAA	8940
CGTACTGcCG ATTGATCGTA CGTTGTTTTT TTGCATTGGC AAAGACATTT ATCAAGTGAT	9000
AAATAACACA AGTAAGCTCA ATTAGTTTAC AAAGTAGCTA TTTCTCtCTA TACTGACTGT	9060
TGGAGGTGAA ATTCATGGTC TTTTACACA TTATTTTGG GATGTTTGCC TTTATTGGCG	9120
CTTTAGGTTT AGGTGTCGCC TTCCGCGATA GCTTTACAGA ACTTTCAACA CC	9172

(2) INFORMATION FOR SEQ ID NO: 66:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 788 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

ATTATTGATC GAGAAATGA AATTCGAGAA TTTAAACCTG AAGAATATTG GAGCATTGAT	60
GGGAATTTCC AAAAAGGTCG CAAGAAATTC AAAGCAAATT TCTGGGGCGT AGATGGTAA	120
AAGAAAAAAT TACCAAATGC AGAAGCGGTG AAAGAAATCA CTTCTCGGAT TGACGGTAA	180
GATTACGATG TTACAAAAGT TGAGAAAAAA GAACGGAAAC GTAATCCTGC ATTACCTTTT	240
ACAACCAGTA GTTTACAACA AGAAGCGGCA CGTAAATTAA ATTTCCGTAC CCGTAAACC	300
ATGATGGTTG CGCAACAATT GTATGAGGGA ATTGCTTTAG GGAAACAAGG CACTGTGGGG	360
CTGATTACCT ACATGCGTAC AGACTCAACC AGAATTGCTG ATTCCGCCAA AGCAGAAGCT	420
GCTGAATTCA TTGAAAAGAC cTACGGAGAT GAGTTTTTCAG CACATGGCGG CCGCAAAGCA	480
AAAAATACGC AAGGTGCACA AGATGCCCAT GAAGCGGTTC GTCCGTCAAG CGTTTTAAGA	540
ACACCTGATG AAATGAAGAA GTATTTAGAT AAAGATCAAT TAAAATTATA TACTCTTATT	600
TGGTCGCGCT TTGTTGCTAG TCAAATGACA CCAGCGATTT TAGATACCAT GAaGGTCACG	660
CTTCAACAAA ACGGCGTAAC GTTTATTGCC AATGGCTCCA AAGTGAAATT TAAAGGGTTT	720
ATGCAAGTAT ATGTTGAAGG ACGCGATGAT GGTAAAGAnG ACAAGGAAAA ATATTTTACC	780
TGAAnTAG	788

(2) INFORMATION FOR SEQ ID NO: 67:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4313 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

TTATGATTAA GGCAACAACG CCCAAGCAC GATATGGAAA TCCCACGCCA CGTGTAGCTG	60
AAACACCTAG TGGTATGCTA AATGCCaTTG GATTGCAGAA TCCAGGTTTA GAAGTGGTTA	120
TGCAAGAAAA ATTACCGAAA TTAGAAAAAT ACCCTAACTT ACCGATTATC GCAAATGTTG	180
CAGGTGCTTG CGAAGAGGAC TATGTGGCTG TTTGCGCAAA AATTGGGCAA GCGCCTAACG	240
TCAAAGCAAT TGAAGTGAAT ATTTCTGTG CCAATGTGAA ACATGGCGGC ATAGCTTTTG	300
GTACAGATCC AGAGGTTGCT TTTCAATTAA CACAAGCAGT CAAAAAAGTC GCCTCTGTTC	360
CGATTTATGT TAAACTATCA CCAAACGTAA CGGACATTGT GCCGATCGCC CAAGCAATTG	420
AAGCAGGGGG CGCAGACGGT TTTTCTATGA TTAATACATT GCTAGGGATG CGCATTGATT	480
TAAAAACGAG AAAACCTATT TTAGCCAATC AAAGTGGCGG ACTATCTgGC CCAGCCATTA	540
AACCTGTTGC TATTCGCTTA ATCCGACAAG TCGCCAGTGT TTCTCAGTTG CCAATTATTG	600
GAATGGGCGG TGTTCAAACA GTGGATGATG TCTTGAAAT GTTCATGGCA GGTGCTAGCG	660
CTGTCGGTGT TGGTACGGCC AACTTTACAG ATCCATATAT TTGTCCGAAG TTGATTGACG	720
GATTACCTAA GCGAATGGAA GAGTTAGGTA TTGAATCGTT GGAACAATTA ATTAAAGAAG	780
TGAGAGAGGG GCAGCAAAAT GCACGATAGA CCAATTATTG CGTTAGATTT TCCCACGCAA	840
AAAGAAGTGG CTGTTTTTTT AGAAAAATTT CCTAAGGAAG AAGCCTTGTT TGTTAAAGTG	900
GGGATGGAAC TTTTTTACGC AGAAGGACCA GCGATTGTTT GCTGGCTTAA AGAACAAGGA	960
CACGATGTTT TCTTGGATTT AAAATTACAT GATATTCCGA ATACTGTGGA AAAAGCGATG	1020
ACAAATTTAG CAAAATTAGG CGTTGCCATT ACTAATGTCC ATGCAGCAGG GGGCGTTCGA	1080
ATGATGCAAG CAGCTAAGGA AGGTCTAATC AAAGGGACTC AGCCAGGCGC TAAGGTTCTT	1140
GAGTTAATTG CTGTAACCCA GTTGACTTCT ACTAGTGAAG AAGAGATGCA CCATGATCAA	1200
TTAATCAACG TTCCTTTAGA GACAAGTGTC ATTCATTATG CTAAATGTGC GGAAAAAGCG	1260
GGTCTGGATG GTGTGGTTTG TTCAGCATTG GAAGCACGTG GTATTCAAGA AGCAACCAAG	1320
CAGACGTTTA TTTGCTTAAC ACCTGGTATT CGTCCAGCAG GAAGTGCAGT AGGTGACCAA	1380
CAACGGGTCG TTACGCCGCA ACATGCTCGC GAAATTGGGT CCACATATAT CGTTGTTGGC	1440
CGACCTATTA CCCAAGCAGA AAATCCTTAT GAAGCCTATC AAGAAATCAA AAAAGACTGG	1500
AGCGAAAAGT AACATGACAA AAGTAGCAAA AAAGATTGCC AAAGATTTAT TAGATATTGA	1560
AGCGGTATTT TTAAACCCTA ATGAGCCGTT TACCTGGGCA AGTGGCATAA AAAGCCCGAT	1620
TTATTGTGAT AACCGTATCA CGATGAGTTA TCCTGCTGTG CGTAAAGAAA TTGCAGAAGG	1680
CTTAGCGGCA AAAATTAAAG AGACTTTCCC AGAAGTCGAA GTGATTGCTG GGACAGCTAC	1740
CGCAGGGATT CCACATGCGG CATGGGTAGC CGATATTTTA GGATTACCAA TGGTTTATAT	1800
CCGCAAGTAA GCCAAAGATC ACGGCAAAGG GAATCAAATT GAAGGCCGAA TTTCTGAAGG	1860
ACAAAAAATG GTAGTTATCG AGGATTTAAT TTCAACAGGC GGTAGCGTGT TAGAAGCGGC	1920

AGAAGCGGCG	GAACGTGAAG	GAGCAACGGT	TTTAGGCGTT	GCGGCAATCT	TCACTTATGA	1980
GTTACCCAAA	GGAACGGCAA	ATTTTGCGAGA	TAAACAGATG	ACGTTACTAA	CGTTAACCAA	2040
TTATTCAACT	TTAATTGATG	CTGCATTAGA	AGCAAATTAT	ATCGAAGAAA	AAGATGTTAC	2100
CTTATTACAA	GAATGGAAAA	AAGATCCTGA	AAaCTGGGGC	AAaTAACCGT	TGAACTAAaG	2160
AAGGTTCTTG	TCGTCTGATT	TTTAGTGTCA	TTTTCAGTAA	CCAGATTTAA	CGGTAGGAGA	2220
ACTTAGGTCA	TTTGCCGAGG	AGCGTAATGA	TCTAAGTTc	TTTGTTTGAC	GTAGAAAAGA	2280
TGTTAGGATA	GAGTTTTAAC	ATGAGTTTTA	CCTGATTACT	ATGTGGAAAT	GrAGGaAGAG	2340
TGTATGAAAA	AAATAAAAAA	TATGGACGTk	GAATGGAGCT	ATACAGGCCA	ACGATGGACC	2400
CAGAACATTG	GCATACGCTA	TGTGACTGGT	TTGCAGAAGG	TGCTAAGTTT	GCTTATCAAT	2460
CGCCAaTTGC	TTTAGAAAAA	GAAaGTGCTG	AGACAGTAAA	TTCTCAGATT	ACGTTTCATT	2520
ATAAAAAAGA	AGAATTCACT	GAAAAAGAAT	TCAAAAATAC	GTTTCATTTT	GTACCACCAA	2580
ATACGGAAAG	TTACGTAATG	TTTGAAAATG	TTGCGTATCA	TTTAACCGAT	ATCCATTTTC	2640
ATATGCCTAG	CGAACATTTG	CTATCGGGAA	AGCAGTATCC	GCTAGAATTC	CATTTGGTTC	2700
ATATGAACGA	TGCTGGCGAA	AATTTAGTTG	TCGTTTGCCT	GTTTACCATC	ACAGAGGAAG	2760
AAAATCGTTT	TTCAGAAGCG	AATCATCCGA	TGGACTGGGA	AAACGGCACG	CACCAACAAT	2820
GGTTTAATCC	GTCAATCTTT	TTACCAGAGG	AGCGTCTACA	CTACCATTAT	GTAGGCTCGT	2880
TAACAACACC	GCCAACTAAG	GGTCCAGTGA	AATGGTTTGT	ATTTGATACC	ATTCAAAAA	2940
TGGATCAAGC	CTTTTTAAAT	AAGATTAAGG	AAGGTATGTT	GGCCTTTAAC	AATCGGCCCC	3000
TTCAGCCGTT	GAATGGGCGT	AAGATTTATT	TTTCAAATGA	TTAGTGTGTG	GAGGAAAAAG	3060
GGTTGAATAT	TCAACAAATG	AAGTATGTTG	CTGCCATAGC	TAACAATGGG	AGTTTTCGGG	3120
AAGCAGCCAA	AAACTATTTT	ATTACACAAC	CTAGCTTGTC	AAATAGTATT	CGAGAATTAG	3180
AAGAAGAACT	AGGGATTTCC	CTCTTTTTTAC	GTACCAATAA	GGGCGCTTTT	TTAACAGAAG	3240
AAGGAATGGT	TTTTTTAGAA	CAAGCAGAGA	AAGTTTTAGT	TCAAATGGAG	TTATTGGAAA	3300
ATCGCTATCG	CGAGACAGTG	ACCAGCGAAC	GATTCTCCAT	CTCTTCGCAA	CATTACGATT	3360
TTTTAGGTGA	GGTAATTGCC	AAAGTTCTTA	AAAAGTATGG	GGATCAATAT	AAAGATTTTC	3420
GTGTCTTTGA	AACCACGACC	TTAAAAGTAA	TTGAAGATGT	CAAAGGATTT	CACAGTGAGC	3480
TGGGGATTAT	TTATTTAAAT	GAACAAAATA	GTGTCAGTAT	TGAACGGTAT	TTAGAACAAG	3540
CTAATCTAGC	CTACGAGGTA	ATTAGTACTT	TTAATACACA	CATTTTTTTA	GGAAATCATC	3600
ATCCTTTAGC	CAAGCAAAAA	GAAATTCATT	TAGAGGAATT	AGTTCCCTAT	CCACAAGTTC	3660
GCTTTAATCA	AGAAGGAAGC	AATTTTTCCT	ATTTTTCTGA	AGATTTAGTT	GAAATTCCTG	3720
AACAGGAATC	GGTCATT CAT	ACTACGGATC	GTGGAACCTT	GATGAATTTG	TTGGTAGAAA	3780
CCAATGCTTA	TGCTTCTGGT	TCAGGAGTCG	TAACGGGTTT	TACGAAGAAA	GAAATTCGGT	3840
TAGTGCCATT	GGCGCCAcTT	TGGAAAATCG	TATTTGTTTA	CTCTTTCCCA	AAAATCGGGA	3900



513

AATCAGTCCA ATTGGTCGCT ACTTTATCAA AGAnTTAAAA GcGTTGTTtA AAAAAgAGGT	3960
AGATACCAAA ACATTTmGCa CTAAGGaTaA AAAGAAGAGg TCGGGACAGA AGTGTTTAAC	4020
TCCGAGAAAT AAGAAGGAAT TTCCGAAAT TGTTCTTTAA TTTTCGGAGA ATTTCCGCTT	4080
ATTTCCGAAG GATTTGCTTC TGTTCCCGCC GTTTATTAGT TTTTGAGCGT GGAGCAAAAA	4140
TCCAAAGTGA TTTTGTCCC ACGCTCTTCT TTTTGGCTC ATTTTGAAAC AGTAGAAATG	4200
GTTkCTCTTT CAATTAATGT GACAGGCACT TGAATGTCTT CTATGATTG AAGGTCTTGT	4260
TGCATCGTTT GTAATAACAC TTCACCAGCT AGTTTTCCCA TTTCTTTAA GTC	4313

(2) INFORMATION FOR SEQ ID NO: 68:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8921 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

ATTnTTnCAT GCGATAAATG GTATTCCTGG TTTATCAGCC GTGAAGCCTA AAGCAGCTTT	60
CTATATTTTT CCGAAATTGA TACGGCCAAA TTTGACTTTT ATGATGATGA AAAATTTGTT	120
TTaGATTCT TACACAAACA TCATATTTTA TTAGTCCATG GTGGTGGGT CAACTGGCAA	180
CAACCCGATC ACTTCCGTAT TGTTTATTTA CCGAAGATGG AAGACTTAAA AACTACAGCA	240
GATAAAATGC GTGAATTTTT AAGTACCTAC AAACAAAAT AAAAGAACTA CCAAAAATA	300
GCCTCTAACG AAGCTATTTT TTGGTAGTTC TTGTTTTATC AACAATACTA TTCTGCGGTT	360
TCACATGCTT TTTTACTAAT AATCATTCGG TTGGTATATT TTTCTTCCAT ATCAGGACTA	420
GGATTGCTA CTACGATGAG ATAAGAATTT TCATACTCCT TCTCCACAAA GCCTGAATAA	480
TTTGTTTTTT CCCAAATGAA GTGAACCTTT TTCTCTTCGG CCATACAATA AACTCCTTT	540
TTAATAAAAA TGAACCTATT TATTTTCTAA GAAACCTTA GATAATGTTC TTGAGTATA	600
TTATGCAATA AAGTACAATA AAAGCATAAT GCGAATTCGC ATATTTTGAC AAAATCCCCA	660
TTTTGCAAGG AAAATGGCCG ATATAAATAG GAGCTATTGT GATGAATTCT TTAGGTGCAG	720
TTATTAAAGA AATTAGAAAA AATAGAAAAT TGACACAGAA AATGTTATCA GAAGACATCT	780
GTTCGCAAAG CGTTTTAAGT CGCATTGAAA ACAATGAAGA GTTGCCAAAC GTTCTCGTAA	840
TGCAACAACT TTGTGATCGT TTAGGTGTGA CTGTTGATCA GATTATGCGT TATAAATCTG	900
GTGATGTGCA CGTTGTTACG TATTCATTG AAAAAATGGC TGAATATTTT CGCCATAAAA	960
AATATCAATT ATTATTAAAT TATTTAAAGG AAAATCGTAT TGAAGAACAA TTGTATTTAG	1020
ATACGGACTG GCAAAAATAT TATTATTATT TAGGtAGTTG TGAATTGTTT GTTCTAAATG	1080
ATTATGAAAA AGCTATTGCT AGCTTGCGTA AAGGGTTATC TTTCACCTAT AAAGCCGATA	1140
AGTTAAATGT TTCTGATTTA GAGATTCAAT TAATTAGTTG TCTAGGTGCC ACTTATGGTT	1200
ACATGGGCAA TCGAGTAGAA GCAGAACGAT TTTTAAGTTT GAGTATTCAT TATTTTAATC	1260

AATTGCCAAA TGAACGCAGT AATGCAGAGT TAACGAAAAT ATTTTTTAAT TATGCTGATT	1320
TCTTATTTAA AAATTATACA GAAAAAGATG CCGAAATTTA TGTGGATCAA GGAATTACTT	1380
GGGCCCCGAA AAAAAATAGT TATTATTATT TAAGTGAATT ATTAAATCTA AAATATTTAG	1440
TTTTAATGAG AAAGAATAAA CACGAAGAAG CAGAACGTTA TCTGAACTTA GCACAACAAA	1500
TGAAAAATGT GGAAAGTGGC AATCTTTAAA AAATCCTCAG AATTGCTTAA AAGGTCTACA	1560
CCATTTCTTG ACTTTTGTAC TGATTAAGAG TACTTTTGAA GTAAGTAAAA TAAAGAAAGA	1620
AGGATGAAGA TGAAAACATT TATCTTTGCA GACAAGTTCT TTTTAAAAAG CGATGTTAAA	1680
GGACCTGGTT ATTTAGAAAT CACGGATGGT ATCTTTGGTA ATTATACAAA AGATGAACCG	1740
CAAGGGgACG TAAAAATTAT TCGTGAGGAA GGCAAATGGA TGCCCCAGG TTTAGTCGAT	1800
ACTCATATTG ATGGTTACAT GAATCATGAC GTAATGGATA ACGATGCAGA AGGCATTAAA	1860
GTGATGTC TG AAGGCTTATT ATCTTGTTGGT GTAAC TTCAT TTTTACCAAC AACGTTAACT	1920
TCAAGTAAAG AACGTTTAAAC AGACGTTGCA CGTACGATTG GTCAAGTCTA CCAAGAAGTA	1980
CCTGGTGCAA AAATTCAAGG GATTTACTTT GAAGGACCTT TCTTTACTGA AGAACATAAA	2040
GGCGCACAAA ATCCAAGTTA TTTTGGAGAT CCAGACCTTG ATACATTCCA CGAATGGCAA	2100
GAAGCTTCAG GCGGAATTAT TAAGAAAATT GCGTTGGCAC CAGAACGTAA TGGTGTGAAA	2160
GAATTTGTTG AAACCGTTAC AGATGAAGGT GTCGTTGTTG CTTTAGGCCA TAGTAATGCG	2220
ACCTTAGAAG AAGCTGATGT GGCAGTTGAA GCAGGCGCTA gCGTTTTTGT TCATGCATAT	2280
AATGGGATGC GCGGCTTAAA TCACCGTGAA CCAGGGATGG TTGGTGCAAT GTTAACGTTA	2340
CAACACGTTT TCTCTGAATT AATTTGTGAT GGACACCATG TACATCCGCA AGCGGCTGAA	2400
GTATTGATGG AAAAAGCGGG CCATGATCAC GTTGCGTTAA TTACAGACTG CATGATGGCT	2460
GGCGGTATGC CAGATGGCAA CTATAATTTA GGTGAATTC CTGTTGTGGT AGCAGAAGGA	2520
ACGGCTCGTT TGGATACTGG GAACTTAGCA GGTAGTATTT TAAAACTAAA aGAaGCAaTT	2580
AAAAATGTAG TTGATTGGGG TATTGCAACA CCTGCACAAG CAATTATGAT GGCTTCATTA	2640
GTTCCtGCGA TTAGCTGTAA AATTGATGAT CAATGTGGCA TGATTGCAA TGGTCGCGAT	2700
GCTGACTTTA TCGTATTAGA ACCAACCATG GAATTAGCAG CCACTTATTT AGATGGCGTA	2760
GAACGTTATC GAGCATAAGA CACAAGAAAA AGCTTGCGGA AATTTATCCG CAAGCTTTTT	2820
TTATAATTCT TTTGGTAAAA TCGCAGAAGT GATTTCTTTG TTCTCGTAGG CAATAAAAAA	2880
ATCGATTAAT TGCTTGTCAA ATTGCTCAAT ATCCTCTTTG TAAAGGGAGC TAGGAATCAC	2940
AAAGGTTTGC TCTAGGGTGC CTGCTAGATA ATCGCCATCT AAAC TTTGGG CGACAATAA	3000
AAATTC TTGT TCGCGCCAAT GTTCCACATC TGTTTCAGAA ATGGTTTCTG GAGAGGCTAA	3060
AAATTTACAA TTTTCGACAA ACTGTTTCATC AATTTCTTCA CTGAAAGTTA ATTCAGCATC	3120
CTCTTCTTGA GGTTGAGAGA GGATTGCTTG CAAAAATGTT TGTGGTTTTG CATCGGGTAA	3180
TGTTTCTTTT AGCTGTTCAA AATCCAATGA AAGTTCCTCC TTATTGTTTG TTTAAATAG	3240

TTACTGAAAA ATCAATGGGT GAATTTGTTT TAAGGCTTTT GGCTTCTAAG ATGATTTTTT	3300
CAGCCAACAA ATCGTTTCCG TTATATACAG GTGTCACGCG ATAGCGAACG GTTTCGCCGC	3360
GATTAAGAGC TGCCCGAATC TGATTTTCAT ATTTCTGCAT ATAAGGAGTG TTGACAGGAT	3420
TCTGATACAA GGTGGTTAAG TTACGTGGAT CATCACCCT GCCTCCCATT TGACGGCCGA	3480
TTAGGTGGCC ACGCGAATGA TTAGCTTTTC CTGAAATGAA ACCAGCTGGT CGAATATCTT	3540
TATTAGCGGA AGTTCCTGTA TTGACCATTG CGGGTTTTAA TAGCGCATCA GCGCCTGTCTG	3600
CTCGTTTTAG ACGATCTAGA GAATGATAAG TAATCCAGCC GTTTGAGTTA TCCTTTAATT	3660
CCTCTGCTGA AAATGTGGCT TTACCATTTT CGATGGGTnC GGGATTATCT CCTTTTGGTG	3720
CTGGTGTGCT TGAAaGGGgT GGTGTGTTGT GGTGAACAT TAAATAAATC TTGTAAAGCA	3780
TCAGGAACCT TAACGCCAAC CGCGCCTAAA ATTAAAATGA TTAATGCGCC AATCAGCATG	3840
ATGGTTGAAT TAAATGGTGG TTTCTTCTTT CTGTTTGCCA TAACGTGCTC CTCTAATTTT	3900
ATACTTGCAA TAGTCCAAAT TTAAGTGGAC CATCTTTTCT ATTATACGTA AATTTGCTAA	3960
AAAATCAAAA AAGGCTTTTC TTTGAAAGAC TTGTTTTCTC AGATATGTTA CTATTTTAAT	4020
GTAGAAGGGa AgGGATAGTG GATGGCCTAT ATTGAAGTCA AAAATGAATA CAAGCGTTAC	4080
CAAATGGGAG AGACAACCTAT TACTGCGAAT GACGGCATT CTTTTGAAGT GGAAAAaGGA	4140
GAAGTAGCGG TTATTTTAGG CCCCAGCGGT GctGGAAAAT CAACTGTTTT AAATATTCTT	4200
GGTGGAATGG ACAGCTGTGA TGAAGGGGAA ATTATTATTG ATGGCACAGA TATCGCTCAG	4260
TTTTCAGAGA AACAATTAAC CACTTATCGA CGAAATGATG TCGGGTTTGT TTTCCAATTT	4320
TACAATTTAG TCCCAATTT AACGGCAAAA GAAAATGTCG AATTGGCTTC GCAAATCGTT	4380
GCGGATGCGT TGGATTTCGAC CAATGTTTTA CAATCGGTGG GTTTAGGGGA ACGTTTAGAT	4440
AATTTTCCAG CGCAACTTTC TGGTGGTGAA CAACAACGAG TAACCATCGC CCGTGCGATT	4500
GCTAAAAAGC CCAAGTTACT TTTATGTGAT GAACCGACAG GAGCGATGGA TTATGAAACA	4560
GGGAAACAAA TTTAACCAT TTTGCAAAAC ACTGCTCGTG AACTGGAAC GACCGTCTTA	4620
ATTATTACCC ACAATTCAGC CATTGCAGAA ATGGCCGATC GGGTAATTCTG AATTAATGAT	4680
GCGAAAGTTC GCGAAATGAC TGTGAATGAT CAGCCTAAAT TGTTTGCTGA AATTGAATGG	4740
TAGGTGCCTA AAATGAAGAA AACAGCGCTA ATTAAAACAA GCTTACGTGA AATTAGACAA	4800
TCTACTACAC GATTTTTATC GATTATGGGG ATTATTTTCC TTGGTGTAAAT GGTCTTTGTT	4860
GGTTTGAAAG CTAAGTGACC AGATATGATT AAAACCGCGA ATAATTATTA CCAGAAAGAG	4920
CAGCTTCCAG ATGCGCGAAT AATTTCTACA ATGGGACTGG AAAAGAAAGA TTTAACAACA	4980
CTACAGTCTT TAAAAGATGT CGAAACAGTT GTGCCGCGGT ATACAATAGA TGCCACGATT	5040
GGGCCGCAAA ATAATGCAGT AAAATTGTTT GGCTATCGAA AAAATCAAGC AGGATCGGTT	5100
AACTATCAAG TAGTTGACGG TCGGTTACCT AAACAACAA ATGAAATTGC ATTAGATACA	5160
TTAGCGAAAA CTAGATATGA CTATAAAATG GGTGATAAAA TCACTTTAAA CGATGCAGCG	5220

ATTAAAGAAA AAGGGCTGAA ACAAACTCAT TTTACTGTTG TTGGATTTAT CAATTCACCA	5280
GAATATATTG ATAATACATC CAGGGGAACC ACAACGGTCG GGTCTGGAAC GTTAAACTTT	5340
TTTGGCGTGG TTTCGGAAAA AGCCTTTGAC AGTCAACGGT ATCCAGAATT GCTAATCTCA	5400
TTTCGTTTCA TAAAACATCA AAATACGTAT TCCTCTGATT ATGAAAAGAA AAGAGAGCAG	5460
GCCTTAAATC AAGTAAAAGA GGCCTGAAA AATCGGCCAG AAGAACAGGT TGCGGCTCTG	5520
CGTGATTCGG CTCAGCCGGA CATTAACCAA GGTGGGCAAG CGCTGGAAAC GGGAAAACAA	5580
GCACTTGCGC AGTTGGaGCA ACAACCAGGA ATTCCAGCTG AGATGCTTGA AAAGCAAAAA	5640
GACGAATTGG CTAAACAAGA ACAACTGTTG GCTGAAAAAG AACAGAGCT AGCCAATTTG	5700
AAAGCCCCaA CGTATTATTA TTTTACTAGG GAAGACAATC CAGGATTTTC TGAATATCaA	5760
GATAACGCTG ATCGGATTTT TTCTTTAGCC ACGCTGTTCC CGTTGTTCTT TTTCTAATT	5820
GCGGCATTAG TCAGCTTAAC TACCATGACA CGTATGGTTG AAGAAAAAAG AATGGAAATT	5880
GGTAGTTTGA AAGCTTTAGG TTATCGTAAT CGTGAAATTG CTTCGATTTT TATTACCTAT	5940
GCAACGGTTG CCAGCTTAAC AGGTGCATTA TTAGGATTGG CAGTTGGCTA TTATCTATTC	6000
CCTAAATAAA TTTTGTATGc CTACGGTCAA ATGTACAATA TTCCTGATTT AGTTACACCG	6060
TGGTACTTAA ATTACAGTTT ATGGGGCATC ATTGTAGCCT TAGCTTGTA TGTGGGAGCG	6120
GCCTTAGTGA CATTAAGAAT CGATTTATTA AGTACCCAG CTACCTTGCT AAGACCCAAA	6180
GCGCCGAAAG CAGGCCAACG GATTTTACTA GAACGAATCC GCCCTTTATG GCAACGCATG	6240
AGCTTTATTC AAAAAGTGAC CGCTAGAAAT CTGTTCCGCT ATAAACGACG GATGCTGATG	6300
ACCGTTATCG GAATAGCTGG TTGTATGGGG CTGTTAATTG TTGGTTTTGG TTTACGCGAT	6360
TCAATTGTGG ATGTCGCTAC AATTCAATTT AATAAAATCT GGCATTACCA AGCAGTCGTG	6420
ACTTTTAAAG AGCAAACAAC ACCTGAAGAG ACAAAGAGT ATCAAGAAAC GTTGCACAAA	6480
ATGGATGGTT TGAATAAAAC AATTCCGTTA TATTCAGAAA TCTTTAAAC GAAGGGCAAA	6540
GGCGCACCAA CTCAAATAT TACTCTTTAT GTGCCACAAG ATCCTTCGAA AATGGCTGAT	6600
TTTCAATTAT TTAATGATCG AGTCACTGGA GAAAAATATT CCCTCAACGA TGACGGTGTA	6660
ATCATTAATG AAAAATTAGC TAAATTATTT GGCTATAAAG TCGGAGATCA ATTAAATTTA	6720
GAAAATAGTG ATAATCAAAC CTATCACGTG AAAATTGCCG CGATTGCAGA GAACTACACG	6780
GGTCATTTTG TCTATATGAC GCCGAACTC TACCAGACGA TGACGAAACA GAAGCCGGAG	6840
TATAATACAG AATTCCTACT TTTTGACAAA AAGTTATCCT CAAAACAAGA AACCTCAATC	6900
GGAGAGGCGT TGACAAAACA GCCGAAAGTG TTGAACATTA CGTTTTTAAC TGCGATGAAA	6960
GGTTCGTTTG ATGATATGCT GAAAAGCCTT GATATTGTCA TTTGGGTACT AATTGCTGTT	7020
TCTGGTCTT TAGCGTTAAT TGTGTTATAT AATTTAACCA ACATTAATGT TTCTGAACGG	7080
ATTCGAGAAC TTTCCACGAT TAAGGTTTTA GGGTTTTACG ATCGTGAAGT CACCACCTAT	7140
GTTTACCGTG AAAACATTAT TTTAACGTTT ATTGGTATTA TTGTTGGCTG CTTCTTTGGG	7200

```

AAAATTTTGC ATCAGTATAT TTTAGCAACC GTTGAAGTTG ATTTAATTAT GTTTTCGCCA 726C
ATTATTCATT GGCCAAGTTA TCTGTATTCC GCAGTCATCA CCATGTGTTT CACGTTGTTT 732C
GTGATGGTCA TTATGCATCG GAAATTGAAG AAAATCAATA TGATTGAAGC GTTAAAAATCA 738C
AATGAATAGC ACATTTTAAA AAaCGAGGAG TCCGTAGAAG TTCGATAATT CTAGGGCTGC 744C
TCGTTTATTG CTGCTTCTTT TATTAAGAAA AAATAGGTCT TAAGATGTAT CTTTGTAGAT 750C
TTATTTGGAA ATTACTACTC TTTTATTTTA AAAATTGGTA CACTATTACT AGGTTATAGC 756C
ACAGAAAGGA GTGTGGGAAA AGTGTTAGAA GGTCTACCAT TGGAAACGGT TTATTTTAT 7620
GCATTAGTGG GGAGCsCgGT CCTCGCTTTT TACTAATTA TTTTGGGGA TGTTTTTAAT 7680
TTTGATGGAC CGGTTGaCCC AATGCTGATT ATTCCATGGA TTGCTTTTAC CTCGCTGTTT 774C
GGGTACCTTG GGAAGAATT AACAGCGGTG AATAGTTGGC TTATTTTAAT TGCAGTGGT 780C
ATTCTTTCAA CCATTATTGT TTTCTTCCTA AATTTTTATG TGTGGTCCC TTTGAAAAAT 786C
TCGGAGGCCA CAATTTCTAT TTCGAAAAA GACATGGAGG GTCGTGTTGC TACTGTGATT 792C
ACCCCAATTC CTGTCCGGGG AATGGGCGAG ATTCAATTGA AGAGTGTCAC AGGTTCGTTA 798C
AGTCGACCAG CAGCATTTTA TGTGCCACAA GAAGTGGCTG CGCCAAGAGG TAGTGAGGTA 804C
TTGATTATTG AATTGAAAGA ACGGGTTTGT TATGTCATTC CTTACGAAGG CAGTCTAAAA 810C
ATATAGAGTA GGAGAGAAGA AAATGGGGAT TTTGTTACCg ATTATTATTG CaGTTTTAGT 816C
ATTATTAATG TTATTAATTG TGTTTGTTTC TAAATATCCA GACAGCCcAA ACCTGATGAA 822C
GCGTTAATTA TCcAGCGGGA GCTATCTAGG CTCTAAAAAT GTTCATGTAG ACGAAGGTGG 828C
CAACAAAATT AAAATCGTTC GTGGCGGTGG TGCGTTTGTC TTACCAGTGT TCCaACGTTC 834C
AAATCGAATT AGTTTGCTTT CAAGTAAATT AGATGTTTCC ACACCAGAAG TGTACACAGA 840C
ACAAGGGGTT CCTGTGATGT GTGATGGGAC ATCAATCATT AAAATTGGTT CGTCAGTTGA 846C
AGAAATTGCG ACAGCAGCGG AACAAATTTT AGGAAAAACA ACGGAAGAAT TAGAAAAATGA 852C
AGCACGTGAA GTATTAGAAG GACATTTACG TTCGATTTTA GGTTCATGA CAGTGGAAGA 858C
AATTTACCAA AATCGTGATT AATTTAGCCA AAGTGtACAA GAAGTTGCCa GTGTTGACTT 864C
AGCTAAAATG GGCTTAGTTA TTGTGTCGTT CACAATTAAA GAAGTTCGTG ATAAAAATGG 870C
ATACTTGGAT TCATTAGGGA AACCAAGAAT CGCTCAAGTT AAACGTGATG CAGATATTGC 876C
AGAAGCAGAA GCCTTGAAAG AAACTCGCAT CAAAAAAGCA GAAGCAGAAA AAGAATCACA 882C
ACAAGCGGAA TTGCAACGTC AAACAGAAAT TGCAGAAGCT TCAAAAGAAA AAGAATTGAA 888C
ATTAGCGTTA TACAAACAAG AaCAAGATAT TGCCAAAGCC A 8921

```

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5523 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

TGCAATGCA CGAGGAGGAA CAAAAGAATG GAACArGGAA CAGTAAATG GTTTAACGCa	60
GAAAAAGGTT TTGGGTTTAT CTCACGCGAA GACGGAAGCG ACGTATTCGT ACACTTCTCA	120
GCTATCCaAG GTGATGGTTT CAAAACTTTA GAAGAAGGTC AAGCAGTGAC TTTTGATGTT	180
GAAGATTCAG ACCGTGGCCC TCAAGCAGTT AACGTAGAAA AAACTAATA AAAAATAATC	240
CAAAAACACC TAACATGCGT TAGGTGTTTT TTTGTTAGAA AAAATAGAGA AGAAAAATGG	300
ACAAGCATTT TTCTTCTCTA TTAAATGAGA GCTTTATTGA TCACTAACTG TGGCAGTCAT	360
AATCATTGGC AGAATCATTG GATGACGTTT TGTGTTGTTCA AACAAGAAAG GCTGTAAAGC	420
GGTCGTCATG GCTTCGCCTA ATTTTGCTTC GGTACAGTTT TTATCTTTCA TGCTTCACG	480
TAGAGCATTG AATAAAAGAC GTTGGCCTTC ATGAATCATA TCACCAGACT CACGCATGTA	540
GACAAAACCA CGGGAAAGAA TGTCTGGACC AGCCATAATT TCTTTATTTT TAATATCAAC	600
TGTTGCTACC GCTAAACTA AACCTTCTTC AGATAAGATA CGACGATCCC GTAAACAAC	660
ATTCCCATA TCACCGACAC CATTTCCATC AACGTACACG TCATTAGCGT TGAAGTGACC	720
AGCAGGGCGT GCGCTATCAG CCGTTAAAGC TAAGACATCG CCGTTGCCCA TGATAAAGCA	780
ATTTTCTTCA GGAACCCCTG TATCTTGAGC GAGCGAAcAT GGATTTTAA CATTCTAAAT	840
TCACCATGAA CAGGCATGAA ATATTTAGGC TTCATCAAAC GAAGCATTAA TTTTGTCTCT	900
TCTTGTCAC CGTGACCAGA AGTATGGATA TTATTGATTT TTCCGTGGAT AACTTCTGCA	960
CCGGCTTCAG ATAATAAGTT GATTAAACGG TTAACGCTTG TCGTATTCCC AGGAATTGGT	1020
GAAGTTGAGA AAACAACCGT ATCGCCTGGT TGAATACTAA TTTGGCGGTG TGTACCGTTA	1080
GCAATCCGAC TAAGCGCGGC CATAGGTTCT CCTTGAGAGC CAGTACATAA AATCATTGTT	1140
TCATTGCTG GCAATTGATT CAGTCCGCC GCATCGACAA ACGTTCCTTT TGGCACTTTG	1200
ATGTAGCCTA AGCGTTCGCC ATTGACAATG GCATTTTCCA TGCTACGACC AAAACAGCA	1260
ATTTTGCGAC CTGTTTTAC GGCTGCATCA GCAGCTTGTT GTAAACGGAA AATGTTTGAA	1320
GCAAACTGG CAAAGATAAT CCGACCATCA ATTTTTTCAA AGATTTTAA AATAGAGGTT	1380
CCAATTGTTT TTTCGGATTT TGTAAAGGTG GGTATTTAG CATTGGTACT GTCTGAAAGA	1440
AGACAAAGGA GCGCTTCTTC ACCTAATTTT GCCATTCTGT GTAAGTTGGC CGGCTCACCA	1500
ACGGGTGTAA AGTCAAACTT AAAGTCACCT GTTGCACAA TGTTACCAGA AGGCGTTTTA	1560
ACCACAACGC CCAAGGCATC AGGAATACTG TGAGTCGTCC GGAAAAAGCT AATCGCTGTT	1620
TTTCTAAAC GAATCACCGT ATCTTCATTA ATTTCTGTGA ATTCCGCTTC TCGCAATAAA	1680
CCATGTTTCT TAGTTTGTT GGTAAATAAA GCCAAGGCTA GTGGTCCCGC ATAGATTGGG	1740
ATATTCGCTT GACGTAAAG ATAAGGAACG CCACCAATAT GGTCTTCGTG ACCGTGGGTA	1800
ATAACGAGTG CTTTGACTTT ATGCAAATTT TGAACGATGT AGCTATAATC TGAATCACG	1860

TAGTCAATCC CAAGTAAATC ATCTTCTGGG AATTTGATCC CAGCATCAAT AATGATAATT	1920
TCATCTTGGA ATTGTACCCC ATATGTGTTT TTTCCAATTT CGCCTAAACC ACCGATTGCA	1980
AAAACGCCAG TTTCGTTATT TTTTATGTTT ACTTTCATGT TAAAACTCCG TTAGAGTGAA	2040
GTCCGCATGT TCTTGCTCGT AAGCAAGATG TTTTTCGTCT AATAATTGAA CGTATTCAAT	2100
ATTGTATGGG GTATTTTCTT CAACTAATTG GCGCACGATT ACATCGCTAT CGGCTTCCAT	2160
GTAAAGAGAT TTTGTGTCTT CTCTTTTTTG GTTTCTGATT TTTGTTTCTT GATAGTATAC	2220
TTTGTAATC ATCGATTCAT TTCTCCTTCA GCTACTAATA GTAGCAATTT ATAGTTTTTT	2280
TCATTTTAAG ACAGAAAATT AAACACAACA AAACGAGGTG CTGATCACCT AAAAAAGAAT	2340
GCTCATCAA AACATCTGTT AATGATGAAA TGAAGCCTCG ACTTTGCTTC ATTCTGGGTA	2400
TTTGATTTTA TGTGTTCTTT TTCAACCGTC CCGTGTGTTG TGAATAACAA CAATTAAGCT	2460
TATTTTATCA TAGAATGATA AGGAAGTATA GAAACACTCA CTGGACTTTT AAAAAAACA	2520
AAC TTCTTGT TTCGCCTTAT TCTTGATCAA GTTCGAGTGC CTTTTTTTTA GGACGCCAAT	2580
ACCAACTTAA GCCAATCACG TGAGCTAAAT AAACAATCGC TAAAGTCGTA GTCATCAACG	2640
TATTTGGAC AATTAAACAC GGAATGGCAA ATAAAAGAAA CATGGAAATA AATAGTTTCC	2700
ATTTACCTTT ACGAGTAATT TTTTCTCAA AGAAGGCTTC TTGGATATAT TTTTGATAAT	2760
AAGAAGAAGC TAAAAATTG TTATAAAGTT TATCTGAAC TACGCATCCAA AAAAAGGCCG	2820
TTAATAAATA AAAACCAGTG GTGGGTAAAA GCGGCAAAAA GATTCCTAAC GTGCCTAAAG	2880
CAAACGTACA GGAACCGAGA AAAATGAAGA ATAATTTTTT CATTCAAAGG GTCCTTTCAG	2940
TTTTTCCCAA TTATGGCACA ATCAAAGGA AGTAGCAATT TTAATCGTAT AAGTAAATCA	3000
TTAAATATAT CTATTTTTTA GAAAATCTAA AAGAGAAAGG ATAGAAAGAG AAAGGCTGTT	3060
TTTCTAGCC AGACAGCTTG TATTACCGAT TTCTTTTTGA TACAGTGGGA GTACTAGAA	3120
ATTTAAAGTA AGATAAGAAG GAGATGATTA ATTGAAGTTA ATGTGGCGTT ACACAaTGCG	3180
TTATAAGAAA TTA CTTTTCG CTGAtTTTAT TTGTGTGTTT GGT TTTATT T GATTGAATT	3240
AGGCTTGCCG ACCATTTTAG CACGAATGAT TGACAAAGGA ATTATTCCTA GAGATATGGA	3300
CTATATTTAT CAGCAAGGGA TTTGGATGGT TGTAATTACG ATTAGTGGCG TTGCAATGAA	3360
TATTTTGCTA GGGTACTTTG GTGCCAGAAT TACAACAAAC ATTGTTCGTG ATATTCTGTA	3420
TGATTTATTT GAAAAGATTC AAAC TTCTC ACATAGTGAA TATGaAAGTA TTGGGGTTTC	3480
TTCTTAATT ACGCGAACTA CCAATGATGC ATACCAAATT ATGCTCTTCA TGGGAAATAT	3540
TTTACGTCTT GGCTTTATGA CGCCAGTGAT GTTTATTGCC AGTCTTTACA TGGTGATGCG	3600
AACGAGTCCG TCGTTAGGAA TGTACGTTTT AGGTGCCTTG CCTTTTCTGC TGCTAGCAGT	3660
TGTCGGGATT GCTCGTTTGT CAGAACCGTT ATCTAAAAAG CAACAAAAGA ACTTAGATGG	3720
AATCAATGGG ATTTTAAGAG AAAATCTTTC TGGATTGCGG GTAATACGCG CATTTGTAA	3780
TGAAAAATTT GAAGAATCTC GTTTTAATAA AGTCAATGAA ACTTACACTA AAAGTTCAAA	3840

520

AAGTCTGTTT CGTTTAATGG CAGCAGCCCA ACCAGGGTTC TTTTCTTAT TTAATATTGT	3900
GATGGTCTTA ATCATTGGA GCGGGACTGT TCAAATCAGC CATGGGGATT TAGAAGTTGG	3960
GAATTTAATT GCTTTTATTG AATATATTTT TCATGCGCTG TTCTCGTTTA TGTATTGTC	4020
CAGTGTCTTT ATGATGTATC CGCGGGCTGC GGTTCGGCT TCACGGATTC AAGAAGCCTT	4080
AGACATGGAA CCAGCTATTC GTGAAGAAGA AGGCGTAACA GAAACAGCTA CTAAAGGCTA	4140
TCTAGAGTTT AAAAATGTAA CCTTTCCTA TCCTGGACAT GCGGAAAGCC CAGTTATTCG	4200
CAATGTAAGT TTTAAGGCGT CACCTGGTGA AACAGTGGCC TTTATTGGGA GCACAGGTAG	4260
TGGGAAATCA ACATTAATTC AATTGATTCC ACGGTTTTAT GATGTATCAG AAGGTGAAAT	4320
TTTAATCGAT GGAGTGAACG TAAAAGAGTA CAAACTTAGT GCATTACGCA ATAAGATTGG	4380
CTATATTCCA CAAAAGCGT TACTTTTTAC TGGTACAATT GCTGATAACC TTCGCTACGG	4440
TAAGGAAGAT GCGACGTTAG AAGAAATGGA ACGAGCAATT GACATTGCTC AAGCCACAGA	4500
ATTTGTTTCG CAAAACCGC AAGGCTATGA TGAACCTCTT TCAGAAGGTG GCACGAACTT	4560
TTCTGGTGGT CAAAACAAC GTTTaGCTAT CGCACGGGCG ATTATTCGCA ATCCGGAAAT	4620
TTATATATTT GATGATAGTT TTTCTGCGTT AGATTATCAA ACAGATGCAA ATTTACGAGC	4680
GCGTCTGAAA AAAGAAACAA CAGAATCTAC TGTTTTAATT GTGGCACAAC GTGTTGGAAC	4740
AATTATGCAT GCGGACCGCA TTGTTGTTTT AAACGAAGGC GACGTGGTCG GAATTGGCAC	4800
ACATCGTGAA TTACTTGAGA CTTGTCCAAT TTACTATGAT ATTGCGGCTT CTCAATTGTC	4860
AGAGGAGGAA TTAGCATGAA ACACGCCTTT TCTTCTATGA AACGAATCGG TCGCTATATT	4920
AAACCGTACC GAGTGACGTT TTATTTAGTT ATTTTATTTA CAATATTAAC CGTTGCCTTT	4980
AATGCAGCGT TGCCTTATTT GACTGGATTA CCGACGACAG AAATTAGCCG TAATATTGCG	5040
GCCGGCGAAT CCATTAATTT TGATTATGTA ATCCAATGTT TAATTTGGAT TTTAGTTGTG	5100
GGAACAGGTT ATTGTGTGGC ACAATTTTTC TCAGGCTTTT TAATGACGAA TGTCGTTCAA	5160
CAATCCATGC GCGATTGCG TCGCGATATT GAAGAAAAA TCAATCGTTT GCCAGTTTCT	5220
TATTTTGATA AGAACCAACA AgGAAATATT TTGTCACGGG TGACGAACGA TGTGGATGCT	5280
GTCAGCAATG CGATGCAACA AAGTTTATC AATATTGTTT CAGCAGTCTT AGGTATTGTG	5340
ATGGCGGTAG TGATGATGTT CTTAATCAAT CCGCTGATGG CGATTTTTTC AGTGATTATG	5400
ATTCCGTTGT CTCTGATTAT TTCCAGAACA ATTGTTAAAA TCTCCAGAA ATATTTCCAA	5460
GGAATGCAAA ATTCCTTAGG AGAATTAAAT GTTATGTCCG GGGCCGAGnT CGAATTTTAT	5520
AAT	5523

(2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5794 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

ACAGTCGATT TCGAGCACCG GCAACTTTTG CTACATCAAT GATAGACGTC ATTTTTTACC	60
TCCAAAGTGG AATGTATTCT TATAAGAAGT ATAGCATACA TTTTTTTGAA AATAGTGCTT	120
GACATGGGAA CATTCCCATA ATTTATCCTA AAGATGTAAA CGATAACAAT TTAATCAAGG	180
AGGAAGAAGA CGGATGACAA GAGGAGCATT AAAAATTGCA ACAATTGGTG GCGGTTCTAG	240
TTATACACCT GAATTAATTG AAGGATACAT TAAAAGAAAA GATGAACTGC CAATTAAGGA	300
AATTTGGTTA GTCGATATTG AAGCTGGTAA AGAAAAATTG GAAATCGTTG GGGCAATGGC	360
GAAgcGCATG ATTAAAGCTG CTGGTTTAGA TTGGGAAGTT CATTTAACAT TGGACCGTGA	420
AGCAGCCTTA AAAGGGGCGG ATTTTGTTC AACACAGTTT CGTGTAGGAT TATTAGATGC	480
TCGCATTAAA GATGAACGGA TTCCTTTGTC CCATGGAGTA TTAGGCCAAG AAACAAATGG	540
CGCGGGAGGA ATGTTTAAAG CATTTGGAAC AATTCCTGTC ATTTTAGCTA TTATTGACGA	600
TATGAAACGC CTTTGCCAG ATGCTTGGTT AGTGAATTTT ACTAATCCAG CTGGAATGGT	660
TACGGAAGCT GCAATTAAAC ATGGTGGTTG GAAAAAACA GTTGGCTTAT GTAATGTGCC	720
GATTGGTCAT AGAAAACAAG CGGCCGAAAT GCTGGGAATT CCAGAAGAAG ACTTATTCTT	780
TAAATTTGCT GGAATTAATC ATTTCCATTG GCACCGCGTT TGGGACAAAG AAGGACAAGA	840
GCGAACACAA GAACTCATTG ATTTAATTTA CGGacCTAAG CAAGAGCAAG AAAGTCATTt	900
GAAAAATATT TTTGATGCGC CGTTCCATTA TGAACAGTTA AAAGATTAG GAATGTTACC	960
ATGTGGTTAT CATCGCTATT ATTATATTGA AGATGAGATG TTGAAGCATT CCATCGAAGA	1020
GTATGAACGG GCGGAAACAA GAGCACAAGT GGTAAAGAA ACGGAAGGTC GTCTGTTTGA	1080
ATTGTATAAA GATCCCAAAC TAGATTACAA ACCAAAAGAA TTAGAAGAAC GGGGCGGCAC	1140
ACACTATAGC GATGCAGCTT GTGAAATGAT TGCATCAATT TATAATGATA AACGAACGGA	1200
TATGGTTGTT TCTACTGAAA ATAATGGGAC AATTACTGAT TTACCGTATG ATTGTGTTGT	1260
AGAAGTATCT GGACCTGTGA CAGCTCATGG CCATGAACCT TACAACTGGG GGGCCTTCC	1320
GCCAGCAGCA CGAGGAATTA TTCAAAATAT GAAAGCTATG GAAGAAACGG TTATTCGAGC	1380
AGCCATCAAT GGTGATTACG GCGCAgCATT gCATGCTTTC ACTATTAATC CATTAGTTCC	1440
TGGCGGCGTC ATGGCGAAGA CCTTGTGGGA CGAATTATTA ATTGCCACA AAGCGCATTT	1500
ACCGAATTTT GCAGATGCGA TTAATAAAAT TGAAGAAACA CAACCAGAAA CAGTCACATA	1560
TGTGGCCGAA TTAATGAAAA GTAATTAAAA AATATCGCAC GCTAAAGATT CGTTTAGTGT	1620
GCGATACTTT TTTTAGAATA GGAAGCAAAG GATTCCGTTT TTTTCTTGAC AGTAAATGA	1680
TAGGTTTGTT GCCGGAACAG CCACTGAATA CCAGAGTAAA TTAAAATTTT AGCAATAATA	1740
AACGTCGAAA GTTTCTGTGC GGAAAGGAAT AATAGCGTTA AAATTAAAAA TAAGACGACA	1800
CAGGTATAAA AAACATGTTG CAATTTTTTA GAATTGATAA TCGTGATTTT TTTGGTATCT	1860
TCAGTAACTT CAATGCGCCG ATAAAATTGG AACCAGCCTT CATAGAACGT TAAAACATCT	1920

CCAGGATCAG CCTCAATTC AATAGGTACA TTAGTAAAAA GTTCCCTTG ATGCACACCA	1980
TTTTTATAAA ATGtAAAATA GCGTAATGCG CCCCAAGCTG ATGTTTTTCG TGTCAATGTA	2040
ATCTTCATGG TTTTCACCTC CGtCTTTATT TTGCCmTwGg TtTCGGTAAA ATTGgTATCC	2100
TACTAAAGTC CGATTCTCTC ATTTTAGGGT CAGAATTGAC AAAAGAAAGC AGAGGTGTTA	2160
ATGTGGGAGT AATTAAACGA GGAGGTATAA TTGTATGTTG GCATCTATCA GAATTGGAT	2220
TGACTTTTTG ACAAATAAA TTGAATAAAT CTTATCAAAA AGTCTGAAAT TAAAATTCAT	2280
GGGTGGCGTG TGTACTTTTA TACGAAACTA CAAGGATAGC GATTAGGCTG TCCTTTTTTG	2340
CGTACCCAAG ATTCTTTCGT TTCAGAAATG AAAGAAGAGG GCAAAATGAA AAAATTAAGT	2400
ATTCATCTGA AAGACGTAAG TATTCACTTT TCGGGAAAAC CAATTCTAGA AATTGATGAA	2460
TTATTTGTAT ATGAAAATGA AAAAATTGGG ATTATTGGTA AAAACGGCGC AGGAAAATCA	2520
ACACTGCTCA ATTTAATTAT GGGTAAGATT CAATCAGATA AAGGAAAAGT TCAAAGATTG	2580
AACGACTTTC ATTACTTGGC ACAAGTAGCT GAAGAAATAA CGAATGAATC AGAGAAAAC	2640
GACAAAAATT GTTTGCTAAA TCAGAAAAAT CAAAAACTAA GTGGTGGCGA AAAAGTCCAA	2700
AAACGCTTAG CAACATTATT TTCAGAGTAT CCAACCGGGG TTATTTTAGA TGAACCAACC	2760
ACACATTTAG ATAAAGAACA TCGTCAGTTG TTAGTGGCAG ATTTAACGTA TTATTATGGG	2820
ACCGTTTTGT TTGTTAGTCA TGATCGCTTT TTCTTGAATC AATTAGCAGA GAAAATTGG	2880
GAAGTTTCGG ATGGACATGT CAAAGAATAT TTAGGGAAC TACGATGCGTA TTGTCGTCAA	2940
AAAGAATTGG AACAGCAAAC ACAATATAAT GTCTATCATC AGTATCAAAA GGAAAAGAAA	3000
AAATTACAGG AATCTTACGC AAAGAAACAA GCACAAGCGC AAAAATCTAG TCATGTTTCA	3060
AAAAAACAAA AACAAAAGCA AATTAAACCT AGTCGTTTAG CTGGTTCTAA ACAAAGAT	3120
ACCGTACAAA AAGCACTCCA AAAACAAGCG AAAGsGwTTA ATGCCAGAAT TGACAGATTA	3180
CCCGATGTTG CGCAAGCAAA ACAAGAGCGA AAAATTATTT TTCCTACTAA TAATCAGTTC	3240
TCTTTATACA ATCCATATCC AATTAGAATT GAAAATTTAA CTTTCGCTTA TGAGAATCGA	3300
ACAATTCTTA ACCAAGTGAA CGTTCAAATT CCTTTAAACG AAAAAATAGC ACTTTGTGGa	3360
AAAAATGGTG CTGGGAAGAG TACCTTCCTT CAACAAATTG AAGCCCGTCA CCCGGCCATT	3420
TACTTTTCTC CCAAAGTCCG GTTAGGAACG TaTCATCAAT TAGATTATCG ATTAaaaaAT	3480
GATGAACCAC TGTGACTTA TTTATTGAAG CGGACAAATT ATTCTGAAAA GATTGTGCGT	3540
TCACTTTTAT ATCGACTAGG TTTTCAACAA GAAAATTTAC AGACAAAGAT ATCTTCTCTA	3600
AGTGGTGGCG AACATGAAAA ATTACTTTGG CGCAATTATT TATAGAACCA AATAATATCA	3660
TCTTATTGGA TGAACCCACC AATTTTCTTG ATTTAGACAC AATCCAAGCA CTTGAAGAAT	3720
TTATAAGTGC CTATCAAGGA ACAGTAATTT TCACTTCGCA TGATGAAACG TTTGTGGAAA	3780
AAGTAGCCAC CCGCACTATT TATTTAGAAA ACGGGAAAAT TATTGATAAA TAaAAAAAGT	3840
rAACAAAAAG TCGCTATTTT AGCGACTTTT TGTTTAGACT TCTTCAATTG GTAAATAGGT	3900

GTATTCTGCA ATTTCTGCGG CATGGGAGCC AGCAACATGT CCAGTGACAA ATGCAGCTGT	3960
AACATTGTAG CCTCCAGTAT AGCCATTAAT ATCTAAAAGT TCACCAGCAA AAAATAAACC	4020
ATTGACTAAT TTGCTCTCCA TTGTTTTAGG GGTCACTTCT TTGAGGGAAA TCCCACCGCC	4080
TGTGACAAAA GATTTTTTCCA AAGGCAATGT TTTCGTAACA GTGAACTGAA AGTCTTTCAG	4140
TAGTTCCACA AAGGAAAGmC GTTGCTTTTC TGTTAGTTGT TTGGCAGGCA CTTCTTCAAA	4200
AGATTTTGTC GGAAACACAT CCAAGGCTAC CGTGACAGGT TGATTACCGT TGCGAGTTAA	4260
TTCTTGGTTA ATAAACTAG AACAGCGGAG CGCGGCAGGT CCTGAAATGC CAAAATGTGT	4320
AAACAGCATA TCCaTTTGAT GATTAACTAA AGGTTTTCTT TTTTGGTTCA AAACAGTTAA	4380
ATTAACATCT TGTAAGAGA GACCTTGCAA CGTTTTATCC AGGATAAAAG GTTCTTCAGA	4440
AATAATAGGT GATTCGGTAG GGTAGAGCGG GCTGATGGTA TGCCCCATTT TTTTGGCTAG	4500
TTTATAGCCA TCACCAGTTG CTCCTGTGGA AGGATAAGTG CGGCCGCCAG TTGTTAATAC	4560
AACACACGGT GCATAAATTT TTTCCAGTTC GGTTCACAG CCAATTATTT GATCGTCTTT	4620
TCGTAGTAAT TTTGTGACCT GTGTTTTTGT AAAAACAGTG ACTCCTAATT CGTTAATGCG	4680
GTTAAATAGC GCATCAACAA TTGACTTCGA TTTATCTGTA ACAGGGAACA TGCGTCCGTG	4740
ATCTTCTTCT TTTAAGTGAA TACCATTGGA TTCAAAAAG TTCATGATAT CATAGTTATC	4800
AAATTGTGAA AATGCGCTGT ATAAAAATTT TCCATTCCCA GGAATAAATG AAATGATTTT	4860
TTCTGCGGGC CGATTATTGG TTACATTACA GCGGCCGCCA CCAGTCATTA ATAATTTTTT	4920
CCCAACACGG CGATTTTTTTT CAATCAATAA TACTTGAGCG CCTGCTTCGG CGGCCGCAAT	4980
CGTGGCCATC ATACCGCTCG TCCCAGCACC GACAATAATT ACATCAAATT TCTTCATAGT	5040
TCTCCTCCAA TTCAATAGGT AgTGTACCAT AATTACGGGA TTTTTTCTGG ATTTAAGGAA	5100
AAAATAAACG TTTGACTATG GTAATTTTTG AATTTTCAGT TAAAATAAAC CTATAACAAT	5160
GAGAGGATGA TTAAATGGAC GCAAAACAAT ATGTTAAAAA TATTCAAGAA AAAATTCACC	5220
AATTAGACCA AGGTCAAACA GAATATTTAC AAGCAGTCGA TGAGTTTTTA CCAACCGTAG	5280
AAGGTTTTTT AGAAAAAAT CCTCAATACA TTGAAGCAAA TGTTTTAGGT GTCTTGATTG	5340
AACCTGAACG TATTTTTCAA TTTCGTGTGC CGTGGCAAGA TGACCAGGGC AATTGGCATG	5400
TGAACAGAGG CTATCGTGTC CAATATAATT CAGCGATTGG ACCATACAAA GGTGGTTTGC	5460
GGTTTCATCC AAGTGTGAAT TTAAGTGTGA TGAAGTTTTT GGCCTTTGAA CAAATTTTTA	5520
AAAATAGTTT AACTGGTTTG CCAATCGGTG GCGGCAAAGG CGGTAGTGAT TTTGATCCAA	5580
AaGGTAAGTC AGATGCTGAA GTGATGCGCT TTTGTCAAAG TTTCATGACC GAACTTyAAA	5640
AACATATTGG TCCAAGCACA GATGTTCCAG CAGGCGATAT TGGCGTAGGC GCTCGTGAAA	5700
TTGGTTATTT GTTTGGCATG TATAAAyGTT GCGCAACTAT GATGCAGGAG TGTTGACTGG	5760
CAAACCTTTA GGATACTGGG GAAGCCAAGC AAGA	5794

(2) INFORMATION FOR SEQ ID NO: 71:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4232 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

ATTCGTATGC CaATTATTT gGCTaTTTGg CgAAAAAgCc AnCnGGAGAA gTCaGrGnAA	60
AATGTGGAAA ATTATAGGAA GCTACTGGGT TTTATGCTAT GATTTTAAAT AAaACAAAAT	120
GAAGGGGCGG GAAACATGT TATTTGGTAC AGCAAAAATG AATCGGGAAA ATCATTTAGA	180
AATCGGTGGT TGTGACACCG TAAAGTTGGC GCAAAAATTC GGGACGCCGT TATTTGTCTA	240
TGATGTCGCT CATATTCGCG CACAAGCAGC TGGTTTTTAA CAAACGTTAA ATCAATTAGG	300
TATCAAAAAT AAAGTGGTTT ATGCAAGTAA AGCATTTAGT TGTTTAGCTA TTTATCAAGT	360
GTAAAGGAA GAAGATATTG CTTGTGATGT TGTCTCTGGT GGAGAACTAT TCACAGCTTT	420
AAAAGGGGGC ATGGAGCCAG CGGAAATTGA GTTCCATGGA AATAATAAAA CCCCTGAAGA	480
ATTACGTTAT GCTTTGGATA ACAAATAGG TACAATTGTC ATTGaTAATT TTTACGAAAT	540
TGACTTGTTG GAAGAATTAT TAGCTACACG AAACCAAACA CAAAAGTTC TTTTTCGAGT	600
GAGTCCTGGC GTGGATGCAG AAACACATGA TTACATTTTG ACAGGCCAAG AAGATTCAAA	660
ATTTGGTTTT GATGTTGCAA GTGGCCAGGC AACAGAAGCA CTCGTTTCGTC TTTTATCAAA	720
CCCCTCTTTT GATGTACAAG GGGTCCATTG CCATATTGGC TCGCAAATTT TTGCTGTGGA	780
AGGCTTTTTA GCTGCTGTTG AAAAAATGTT CACTATTTTA GAAGACTGGC GGCAAGTCCA	840
TCAATTTACA GCACGTGTTT TAAATATGGG TGGTGGTTTT GGTGTGCAGT ACACGCAGCA	900
AGATGAGCCA TTAGCACCAG CAACATTTGT TGAAAAAATT GTTTATTCCT TGAAAGGTCA	960
TTGTGAACAA tTAGGGTATC CaCTGCCTGA GCTTTGGATT GAACCAGGCC GCAGTTTAAT	1020
TGCCGAAGCG GGCACAACGA TATATACAGT TGGCGCGCAA AAAGAGGTTT CTGGTGTTTCG	1080
TCATTTTGTT TCCGTAGATG GTGGCATGGG TGATAATATA CGGCCTGCTT TGTATCAAGC	1140
AGTTTATGAT GGTTTTCTAG CGAATCGTGA TGGGCATGAT TCAGTCAAGG AAGTCACAGT	1200
GGTTGGTAAA TATTGCGAAT CAGGAGATGT CTTACTTCGA GATATTTTAT TGCCAGAAGT	1260
TAAAGCAGGG GATTTATTGG CCATTAGCAG TACGGGGGCT TACGGTTATT CAATGGCAAG	1320
TAATTACAAC CGAAATCCTC GACCAGCGgt GGTTTTGTG GAAGATGGAC AAGCAAAATT	1380
AGTTGCTCGC CGTGAGACAT ATGAAGATAT GACAACGTTG GaTTGTTGAT AGATTaAAAG	1440
TAGCCGCCCTC TTTTGGAGAC gGCTGTTTTT TTTCcAAAA ATAACmAAAT GTAAGGGTGA	1500
TTTCTTTCAG AAaTGaAAAA TCCaTAGTAT ATTAAATGAA GAAGGAAATA CTTTAGAAGA	1560
GGGTGTGATT TGCAGATGAA ATGGCTCCTC GGGTTACTGA TCATTTTTCG AATAATTGGC	1620
GTTGTTTTTT CTTTTTACAT TGCCAATTAT TTTTGA AAAA TTGCGTTGCT TAAAGACAAC	1680

CCTTGGTATC ATAAAAAAGG GCATCGGTTA TTAAATCCTG ATAATTTTCA AGAACGGGAA	1740
ACAAGATATA CAAAAATAGA AGAACAACAA AAACAAGAAG GCGAAGCTTT TTGGACAGAG	1800
TCTTTTGCTG AAGATCGTTG GTTGAAAATT AAAGATGAAA CATTGTATGC ACGGTGCTTT	1860
ATTCCTTATC CAGATAACCA TCGTTGGGCG ATTTGTGTGC ATGGTTATCG TTCAAACGGA	1920
AAACGCGATA TGGCATATAC TGCATTACGT TTTGCAGAAG AAGGCTACAA TGTTTTAGTT	1980
CCCGATTGTC GTGCGCACGG GAAAAGTTCA GGAATAAAAA TTGGGATGGG ATGGCTGGAT	2040
CGGTTAGATT TATTAAGTTG GCTTAGTGAA GTTCTTGCCA TTGATATGGA AGCAGAAATT	2100
ATTTTAGTGG GCGGTTCCAT GGGAGCGGCT ACTGTCATGA TGGCAAGTGG TGAAAAGTTA	2160
CCAACGAATG TCCGCGGATT AATAGTGGAT TGTGGCTACA CTTCTGTTTA TGATGAATTT	2220
AAATACGTCC TACATGAATC TTTTCATTTA CCCGCTTTTC CTATTTTAAC AATTGCTAAT	2280
CAATTAGCAT TAAATAACTA TGGCTTTCAA TTGAAAACCG CTTCTTCCGT GAGACAATTA	2340
CACAAAATA CGTTGCCAC ATTCTTTATT CACGGTACAG GTGATCGCTT TGTTCCAATG	2400
ACAATGTTTG AAGAAAATTT AGCGGCAACT CAGGGGATAA AAAAAGGCCT CATTGTTGCT	2460
AAAGCACCAC ATTTATCTTC AAGTGTCTAC GAACCTGAAA ATTATTACAG TAGTATTTTT	2520
GAATTTTGG AAGAAAATG tCCTGCGGTA AAAACAATAT CCGACTAAAA AAGACATGGA	2580
TTTTCATGAT TCCATGTCTT TTTTTGTCT GACTCGTTTA GTTAAATTTA AAGTGCTTAA	2640
AAAGAAAAAT AACGGCGCCA AACTACCAAG GCAAAGAAAC CATTTCATTG AATAAGTTAA	2700
CTCTTGCTGT AAAAGAGATT CTCGTATATA GCTGAAACCT TCGCCATGGA CCCGCAAATA	2760
AATATAAATG TAAGCAAAAA GAATCACTAA AGGCAAAATA CAGACTTCTA ATATTTCCAA	2820
AAGATAAGGA AATGTTTTAT TTAATAATTT GAGGCGTGTC TTATTCAATA GATAAGCAAA	2880
AATGAAAATT GACCATAAAA AACCACAAAA ACTAATCAAA TAAAAAATCA GATTGCGACT	2940
TCCTGCATTC ATAAAAGCTT TCGGGAAG TAGTCTGCCG ATGAATGTAT TTTAATTGT	3000
AAAAGGCACA TCAAAAATGG AAATATAAAG ACAAAAAAGT ACTAATGTTA ACAAGGCATG	3060
CCAAATCTTT TTCTCCATTT CATTTAAAA ATATCTAATT ACTTGGTACA TTCCATAAA	3120
CTGTCGCTCC GTGTTTGAAA TAATCGTAGT TTTGACTATA CCATAGCTGT TCTCAATTTT	3180
CTTGAAAAAA ATACACGCAA CCAGCCTCGA CTATACGAAA CACACAGAGT GTCTAAAAAA	3240
TAAGTTGCCG ATCGCATTTT TTCGCATGCT AGTGTGTCAC TTCTGTTTCG TTGGTGTTTT	3300
CTAATTGCTG ATTTGTTGAT TAAAAAGGG AAGGAGGGAG CGTTTTTCTA GTATAATAAT	3360
ATAAAGAAT ATAGAGAAAA TCGTAGGATG GGAGTTATTC AAAGTGGCGG AAATTACAGT	3420
TAAGACAATT TCAGAGAAAG ATGATTTAAA GAAGTTGTTT TTGGGTCAAG AGTATGACCA	3480
ACAGCAGCCA GTTAAAAATA TTTTAAAGC TAGAACATCG AATGTCCATA TTGAAAAGTA	3540
CGAGGTTGTG GAGGATGTTT TTTTAAAGTT AACTGGGAAG GCGACCGCGC ACAAAGATAT	3600
TATGAACTCA TTTTGGACAA CGTATAAAAT TATGTTACAA CTTGTTTATC CCGATTCTT	3660

TAGACCAGCG GAGGTAATTG GGGAAAATCA AGAAAGCTAT TTAGAAAAAC CTTCTGAACA	3720
GAATTTTCTT GCAGTAAATA GCAAGTATCC GCCTCATGAT TCCGGACCGA TATTAACTT	3780
ATTTTGACCA AGTTTTTCCT GACTATTTAC CAAATCCTGT ACCAAAAAAA TATACATGGA	3840
ATGAATTCT TCTCGACAAT TTTACTAAaT TTGAGCGAGT TCATCmGGAC CCACAaTTAA	3900
aGCGTTTTGC CgAACTTACC CaTTCAATTG GAAACaTTAC AGTAGTTCCA CTAGGATTTA	3960
ACAGCGGAAG AAGTTTATCG TTTAAAGATT ATTGGGATTA TTCTTTAGAA CAACTTTCTGA	4020
TATTTTtagc TTCATTTCAT TCATGGGAAA GCTACGTTCA TACGTATGAG ATGCAACCAT	4080
TTTTAAATGA ACAGTATCAA CCAGTAGCGC TATGGAAAAA TCATTGAAA AAAGACTCGT	4140
TTATCTTACC TCAGAATATA GAGGAGATAA ATGAATATTT AGTTCAAGTA AATCAACGAA	4200
TAGAAAAAAG AGGGCAGCGA ATAGTGAATC GG	4232

## (2) INFORMATION FOR SEQ ID NO: 72:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24004 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

TGGAnTTATG GTTGGGTTTT nAGGTTGGCC GnAGGTTGGG GCCAAAGGTT ACCCGTTATT	60
TTCCTTGGAC CAAnCCACTT TATTTCGGAC CAAAACCCA TTTTAAGAAA CCATTTGTAC	120
CCAAATGGTA AAGCAGAAGA TTTAGATGCT GCTTCCGCAG AATATGAAAA AATTATCGAT	180
GCACATCCAG TAGATATCCA AATCTTAGGT ATTGGTCAAA ATGGACACAT TGGTTTCAAC	240
GAACCAGGCA CACCTCTAGA TAGCTTAACA CATGTCGTG AGTTGACTGA ATCAACAATC	300
AACGCAAACA AACGTTATTT TGACAAAGTA GAAGATGTCC CAACACGTGC TGTTCCTATG	360
GGGATTGGCT CAATCATGAA AGGTAAAAAA ATGATTTTAA TGGCTTATGG CGAAgCAAAA	420
GCAGAGGCCA TTAAAGGAAT GATTGACGGA CCAGTAACAA CAGACATGCC AGCAAGTGCA	480
TTACAAAACC ATCAAGATGT TGTGTAAATC ATTGATGATG CAGCTGCAAG TAAATTATAA	540
TCGTTGATGA AAAAGAACTT GAAACATAAG AATTCCTTA TGTTCaAGT TCTTTTTTTG	600
TTATACTGGA AGTAATAGAA GAAGAGGTGA AAAGGATGTT ACAGCTAACA ATTCCAGAAA	660
TTATTTTACG GTTAGTCCTA GCGATGGTGA TGGGTGGCGC AATTGGTTTT GAAAGACAAT	720
ATAAGAATCG CCCAGCAGGG ATGCGCACTC ACATTTTAGT TTGTATGGGC GCGGCTATTA	780
TTGCGTTAAT TCAGTCGCAA ATTGCAATCA ATGCATTGAA AGATGCGTTA gcTAATCCGG	840
CTTTGACGGG GGTCAATCGT TCAGATCAAG CAAGGTTGAT CGCACAAgTT GTAAGTGGA	900
TCGGCTTTTT AGGTGCTGGA ACCATCATTG TAACGAAACA AAGTGTcACC GGATTGACGA	960
CAGCAGCGTC ACTATGGACC GTTGCTGGTT TAGGGATCGC GGTGGGCATG GGCTACTATA	1020
CGATTGCCTT AACAAGTTTT ATCGGTGTAT TATTTCCTT AACCATTGTC AAACGAATTA	1080

TTCACGTACC AACAAATCAAA AAATTAGAAA TTCGCTATCA ACATAAACAA GAAACCAAAG	1140
AATTTATTAA TCAGTATTTT GAAGAAAaAC ATATTGAAAT TTATGATGTG AACTTCAATG	1200
TTTCTTGGGT GGATGATACA AAAATCTATA CGAATATCTA TACTATTGAT TTGCCTAAAG	1260
GCCTAACGTA TGCCGATGTT ATCGAAGATC TATCTGTTTC AAATAATGTA ACAAATTAC	1320
GATgATTAAT GTGTAGAAAC ATCTATTaTA ATTAATGACA ACCGCTATAA GTTAAAATGC	1380
TTATGGCGGT TTTCTTCTAG GACCTAGGAC TGATTTTTTA AGTTTAATAT TTCAAAAAGA	1440
TTACAAAAGA GTTATGATTG TATTACGAGT TAGAAAGAAA GGTGATAAGT AAGATGAATA	1500
AAAAAATAA ATTTCTTAGT ATTATTGTTG TTGTCATTGT GATTATTGGT GCATTTGTGA	1560
TGAAAGCTTA TAATCAATTG ATTACTTTAG ATAATAAAGT CGAAGCAGAA TGGTCACAAG	1620
TAGAGAATGT AATGCAAAGA CGTGCTGATT TAATTCCTAA TCTTGTCAGT AGTGTTCAAG	1680
GCAGTATGAC TCAAGAAAAA GAAGTTTAA AAGAAATAAC GGAAGCACGT AAAGCATACG	1740
CAGGAGCAAA ATCTTCTTCT GAAAAAGGGC AAGCAAATGA GCAAATTGAA AAAGGATTAG	1800
GTAATTTTGT TACAGTATTG AATGAAGATT ACCCAAAGTT AGCTTCTTCT GAAAATGTCA	1860
AAACCTTAAT GACTCAATTA GAAGGATCTG AAAATCGAAT TTCTGTTGAA CGCCGTAAC	1920
ATATACAAGC CGTTGAAACA TATAACCAAC AGATTTCAAA ATTCCCTGAT AAAATAATTG	1980
CAAACTTTT AGGATTTGAG AAGAAAATA ACTATGTTGC GACAGAGCAG GGAAAAGAAG	2040
TGCCGCAAGT AAAATTTAAA TAAAAATTTA TAGAAAATTT GGGTTAAAAA GAAAGAAGTG	2100
AGGAAGGTTG GATAAAAAAG TCGAAAAAT TCAAGAAAAA GTTCGTTTGT TTTATAAAAA	2160
GCATCAGCGG AATACAAGCA TTGTTGTGGT AATTAGTTTA AGTATGATTT TAATATTAGG	2220
AATGGCTGCT TGGTATGAAG GAAAGGGAAT CTCTTATGAA GCAAAACAAA CTGCACTGAG	2280
AGAGAAGCAA CATGCCTTAC AAAAAAGAG AGACGATATC TTAGCAGGGA AAATGACTGA	2340
AGCGTCAACC CAGACAACAA TAAATAACTT TTCTAAACAA GAAATAACTG GTGAAATGA	2400
ATTAGTGACA ATTACTGGAA CGAATGGGCA GTTGAACATT AATGATGCTA ATATCTATGT	2460
TTCAGATAAT GCAGGTATTG TCTCAAATCA ACTTAAACAG AAAATGTTTC AATTGAATCA	2520
ACAATTGTTG GAAAATGCGA ATGGCGCACA ATTTATGCTT ATTACCGTTC CTGCGTTACC	2580
TAGTGGGGAA TCTGTTGAAT CTTACAGCAA TAAGATTGCT AATCAATTAG GTGTAGGGGA	2640
TCGTGAAAAA AATAATGGTG TGGTTTTTCT AATGGCAATT GAGGATCGCG AGTCTCGCTT	2700
GGAAGTTGGA TATGGCTTAG AGAGTATTTT AACGGATAGC TATGCGGATG ATATTATTAA	2760
TAATGAAGAA GTAAAAGAGG CTTTATAGAGA CGAGGATTAT AATACAGGTC TCAATAAAAT	2820
TATTGATCAA GTATCGGCTG CAATCAATTC AAAAACCACA CAAGTGGATA ATGAATTAAC	2880
AAATATACAA ACAGAGTTAA AACTAACAC AACGAAAAGA AATGTTCTTT TAATGGCAAG	2940
TATTGCTGGG ATGCTCATTT GTGCTGTTTA TATTTTACAA ATTTTAAGAA CAAGAAACT	3000
TGTTAAAAAA ATGTATCAAG ATTATTTGAA TTGTTTACCA ACAAAGGCTG TCCTAAATAA	3060

CTCAGAACAG ACAAAGAAAG TATTAAATAA GATGAAAAAG ACTTCTTTTT ATTGTTTATA	3120
TCTTAATGGA GCAACGTTAT TTGCCACGAA AGGGAAAATA AGACGTGCCA CTAAGCGTGG	3180
TTCTTTACTA AGTATATATC CAAATGCAAA AAAACAATCT TTTGGGCGTT TATTAGTAGG	3240
AGATACTTTA TATAGTTATG ATGGTAGTGT TTTGACGTAT GCTTATTTAA ACTCACAATA	3300
CAATCCGTCT AATCATAGTA GTTCTGGTAG CGGCAGTGGT GGTTCAATCG GTGGCGGCTC	3360
ATTCGGCGGT GCGGCGCCT CAGGAAGTTG GTAAAATAAA AAGCTAGAAA CGTTGATCGT	3420
TTCTAGCTTT TTGCTATTTA TCTAAACCGT AAAGATAGTC ATCATCTTTC ATGGCTTCAA	3480
CTGTTCCTAG TAAATAGCCA TTTCTACTT GGGAGAAGAA GTCATGGTTG CTAGTTCCTG	3540
TCGAAATACC ATTCATTACA ATCGGGTTCA CATCATGGC TGTATCCGCA AATAAGGGAT	3600
CCATACCTAA GTTCATTAAA GCTTTGTTGG CGTTGTAACG TAAAAAGGTT TTTACTTCTT	3660
CGGTCCAACC TAAGTCATCA TATAATTCTT CTGTATAACG TTCTTCGTTT TCGTACAATT	3720
CGTAAAGTAA ATTATACATC CAGTCTTTTA GTTTATCTTG CTCATCTTCT GGTAATTCAT	3780
TAAATCCTAA TTGGAATTTA TAGCCAATAT ACGTGCCGTG CACTGATTCA TCACGGATAA	3840
TCAATTTAAT AATTTACGCG ACGTTCGCTA GTTTATTGTT TCCTAAATAG TAAAGAGGAG	3900
TGTAACCC AGAATAGAAT AAAAACGTTT CTAAAAATAC ACTAGCAATT TTTTCTCTA	3960
ACGGAGTACC GTTTTTATAA ATTTCTGTAA TCCGTTCTGC TTTATATTGC AGGTGTTGT	4020
TGGTATTTGT CCACTCAAAA ATATCATCAA TCTCTTTTTT TGTATTTAAT GTACTGAAGA	4080
TAGAAGAATA ACTTTTGGCA TGGACAGATT CCATGAATTG GATGTTATTT AAGACGGCTT	4140
CTTCATGAGG CGTCCGCACA TCATTTCTCA GTTGATCCAT GCCGCTTCT GATTGAACAG	4200
TATCTAATAA AGTTAAACCG CCGAAAACGT AGCCGACAGT AGTTTTTCT AAATCGGAAA	4260
GAGTCCGCCA ATCATCTAAG TCATTGGATA ACGGAATCCG TGTATCTAAC CAAAATTGTT	4320
CGGTTAATTT TTCCAAGTG GATTTATCAA TGACGTCTC AATGGCGTTC CAGTTAATTG	4380
CTTCATAATA CGTTGCCATG GTTATTCCTC CTGTATGCTG GGACAGCAGT CTGTTTGA	4440
AGTTGTCTCG TTTCTTCGCT TAAAACGTTT TACAGAGTCC AAATAGGAGT GTGCAGCCGT	4500
TCTTCTTTGT CCAGTATATC TGAAAATGTC CTGTTTAAAA AGAGAACATT CCTAAAAAT	4560
AAGACGCTCA AGCAGAACGC CTTATTTAAG TGCTTTTCAT TAAATCACGC AAGACTCACA	4620
TTGATTGCTG CCAATTTCTT CAGCATCATC TGTAAATGTA CGAATATAGT AAATTGATTT	4680
AATCCCTTTG TTGAACGCAT AATGACGTAA AATGTTCAAG TCACGGGTTG TTTGTTTTGG	4740
TGTTTCTTTC CATTCATACA AGCCTTCAGG GATTTCTGAA CGCATAAACA AGGTTAACT	4800
CATTCCTTGA TCAACGTGTT GTTGCCTGT TGCCTAGACA TCAATCACTT TGCGCATATC	4860
CATGTCGTAA GCAGAAGTAT AGTAAGGGAT TGTATCATTA GCTAAATAAG GCGCTGGATA	4920
GTAAATCTTA CCAATTTTCT TCTTTGACG TTCTTCAATC ATACGAGTAA TTGGATGAAT	4980
GCTGGCACTA GTATCATTA TATATGAAAT TGAACCATTT GCGCTACGG CTAAACGATT	5040



TTGATGATAA AGACCATCTT TTTTCACAGC ATCGCGTAAC GCAGCCCAAT CTCAGCCGT	5100
TGGAATCTGA ATGTCTTTGA AAATTTCTTT CACTTTATCA AATTTTGGTG TAAAGTTGCC	5160
TTCAATATAT TTATCGAAGT AAGCACCAGA AGCATAATCT GATTTTTCOA AATTATGGAA	5220
GACTTGGTTA CGTTCTTTTCG CAATTTGGTT ACTTTCCATT AAAGTCCAGT AATTTAATAA	5280
CATGAAATAA ATGTCTGTAA AGTCTAATGA TTCTTCAGAG CCGTATTCCA TATGATTTTT	5340
AGCGAAGAAG GTGTGTAAGC CCATTGCGCC TAGTCCGATT GTATGGCTTA AACGATTTC	5400
ATTTTGGATA GTTGGCACGA CATCAATTTT AGAAGCATCT GTTACAAAAG TTAACGCACG	5460
CGTCATAGCG CGGACAGATT TTCCAAAGTC AGGACTTTCC ATTAAATTAA CAATGTTAGT	5520
TGAACCTAAG TTACAATAA TATCTGTTCC TAAACTTCA TATTCTTGTT TCCCATTAA	5580
AACAGAAGGC GTTTGAACTT GAAGAATTTT AGAACATAAG TTAATCATGA TAATCTTGCC	5640
ATCAATCGGA TTGTTTTTGT TAGCTGTATC AATATTGATG ATATAAGGAT AGCCAGATTC	5700
TTGTTGTAAT TTAGAAATTT CATTTTCTAA GTCACGGGCT TTAATTTTAC GTTTACGTAT	5760
GTGTTGATTG GCCACTAGGT TATCATATTC TTTGGTAATA TCTACATAGG AATAAGGCAC	5820
ACCATATTCA CGTTCCACAC TATAAGGACT AAATAAATAC ATATCTTCGT GTTACGTGT	5880
TAATTCGTAA AATTTATCTG GAACAATAAC ACCAAGAGAC AACGTTTTTA CACGGATTTT	5940
TTTCATCAGC TTTTCTTTTT TAGCAGATAA GAACATTTCA ATATCTGGAT GGAAGACATT	6000
TAGATAAACC ACACCAGCAC CTTGACGTTG ACCTAGTTGG TTAGAGTAGC TAAAGCTATC	6060
TTCAAATAAT TTCATAACAG GGACAACACC ACTTGCTGCG CCATCATAAC CTTTGATTGG	6120
TGCGCCCGCT TCACGTAAGT TGGAAAGAGT AATTCCTACG CCGCCGCCGA TTCGTGATAA	6180
TTGAAGGGCA GAATTAATTG AACGCCCAAT GCTGTTTATA TCATCTGTTA CTGAACTAA	6240
GAAACAAGAA ACCAATTCAC CTCGACGTTT ACGCCAGCG TTTAAGAAAG AAGGTGTGGC	6300
TGGTTGATAC CGTTGATGGA TCATTTATC CGCCAATGTT AAAGCCAACT CTTCATCGCC	6360
ATCCGCAAAA TACAAGGCAT TAAAGGCAAC CCGATCTTCA TATGTTTCTA AATACTCTGT	6420
TCCGGCGTTG TTCTTTAAAG CGTATTGAGA ATAAAATTTA TAAGCCGCCA TGAAAGACTT	6480
GAATTGGAAA TGTTGTTCAT CTAAGAAAGC ATATAATTTT TCAATAAAGT CCATTGAATA	6540
TTTTTCGATG AATTCAGTTT CAATAAAATC TTCTTTTATT AAAAGTTGA TTTTTTCTTC	6600
AGCAGAGCTG AATTTACGTG TATTCGGTTC TACATTTTCT TTAAAGAAGG CTGCGAGTGC	6660
TTCTTTATCT TTATTTAAAG GGATTTGACC ATCAACAGGA CGGTTAATCT CGTTATTCAG	6720
TTTGAAATAG CTGACATCTT TAATTTCTTT TAGACTCAAT TTCTTCCACT ACTTTCTTAA	6780
AGGTCTCCAC GTCTTCGTCT GTTCCGCTAA ACTCAAAGGA GTATAGAAGC GGAACGTGAT	6840
AATCGCGGGC AATATCTTTG GCGGTATAGA CAAATAATTC AGCAAAATTA CGGTTGCCCC	6900
CACCAGCTAC ACCAACTAAC AGTTCTTGGT TACTTTTATA ATCAAGAAAG TCATTCACATA	6960
CCTCTGTAAT TTCTGCATCA TAGGTAGGAA TGACAAGAAT AAAAGGCTCA TTTATTTCAA	7020

AAAAAGGATT CGCTGGCTCA AGTTCATAAG CAGGCAAATC CAATTTTTTA ATAAAGCGCC	7080
TTGTTTGACC GGTGACTGAA AAATAAACAA GTTTCATCTT AGCTAGCCAA TTGTTTAAAT	7140
TGGTCTGGAC GGAAGCCAAC AACGGTTGTT GCTTCAGATG TGATTACCGG TACGCTTTGG	7200
AACCCTTGTT CTTTTAACCA GTCGATAGCA TCTGGTTGTG CGTCGATATT GATTCTTCA	7260
AAAGCGATGT TATTTTCACT TAGGaaACGT TTCGCCATTT TACATTGGAT ACAGTTATTT	7320
TTTGAAAAGA TTTTACGTT CATTGaAAGT TCCTCCTCTT TTTTCTACA TAGACTAGTA	7380
TAAATCTATG TAGAAAaaAG TCAACACTTT AACACTATAT ATTGTGGTGC GTTTTCTCA	7440
TAATACTAGT TTTTGTGTTT TTGAAAGAAA AATGAAACA CGTGAATCCT CGTTTTTAGG	7500
AGAAAAACA CATTGGTCAG TGTTTGACAA ACAAAAAAAT TAAAATTTTG AAAGAGAACA	7560
TTTAGTGAAA AAACACTAAA TGTTGTGTTG ACATTTGAGT TAGGCACAAG AAGAAAAAAT	7620
AAGCGGATTT GTTATTGAAT GAGTGGGATT TTACAGTAGC TAGTAACATA AAAACAAGT	7680
GCTGACTTTT TTTCAAAAAT AGTAACCTGA AAAAAGTGTC TCTATATAGA AGGAAAATAG	7740
AAGAAAACAG GAGATACATC CAGAAATTCa TTGACTTTCC GTGTAAGCAA GCGTATTCTT	7800
TTAATTGAGA ACGATTCTCA TTTTCGTTTG AAACTTTTTC ACATGGAGGA AAGTAAATGA	7860
TTACGTTAGC GCAGGCAGAA GTTGGGAAAG TCTATACAGT CGAAAGCGTC CAAGCAGATA	7920
TTCAGACAAA GAAACATCTG AACAATTTAG GTGTCGTTGC AGGCCAAGCA GTTGTCTTG	7980
TCAATTATCA GAATCAAAAT GGTATTGTTT TATTACATAA TAGTCGAATT GCGCTGACAG	8040
ATACGATTTT ACAAGCGATT CATGTGGAGG AGCGCTCGGC GAATGAGAAA GTTTGGGTCT	8100
CCTTGATAC CTTAAAGTT GGAGAACGCG CTACAATAGT AGGGATCCAC GGACAAGGAG	8160
CGGTGAAACG TCGGTTGATG GACATGGGCT TAACAAAAGG AACAGTAATT TTTATTCGCA	8220
AAGTAGCTCC TTTAGCTGAT CCAATTGAAA TCAATGTACG TGGcTACGAA CTAACGTTAC	8280
GAAAAAGTGA AGCGGaACTA ATCTTGTTG AAAAGGAGGA AAGCGAATGA CTCAGCAAAT	8340
TGCACTAGCC GGCAACCCTA ATAGCGGTAA AACGACAACG TTCAATATGT TAACAGGAGC	8400
CAACCAATAT GTTGAAACT GGCCAGGGT GACTGTGGAA CGCAAAGAGG GGGTGGCAA	8460
GAAAGATAAG AACTCATTa TTCAAGATTT GCCAGGGATT TACTCTCTT CTCCTTATAC	8520
GCCTGAAGAA ATTGTGGCAC GAGATTATCT TTTGGAGGAC CAACCTTCAG TGATTTTGAA	8580
TATTCTCGAT GTGACCAATT TGGAACGGAA TCTTTATTTA ACAACACAAT TAATTGAAAC	8640
TGGATTACCG GTTGTGTTG CGTTAAATAT GATGGATTG TTGGAAAAA ACGGCCAAAC	8700
ATTGAATAGT GAAAAGTTAA GTTATGGCTT AGGCGTGCCT GTTGTGAGA TTAGTGCTTT	8760
GAAGAATCGT GGGTTGGATC ATGCTTTAAA ACAAAGTAAG CAGCGAGCGA ATGCCATAGA	8820
GACAGAGGTT ATTTATCCTT CTTATGATAA TCGCTTAGAA GCAGCACTTG CTGAAATTGT	8880
TGATATTTTA GGAATACTG TTCCTGAAAC GCAACAGCGT TGGTATAGTT TAAACTTTT	8940
TGAACGAGAT GTTCGAACCA AAGAGCAATT ATTATTGAGT TCTTTTCAAG AAAAAGAAAT	9000

TGAAGAAGTC ATTCAAATTA CTGAAAAAAT CTTTCAAGAT GAAAGTGAAG CGATTATCAT	9060
TAATGAACGC TATGCCTTTA TTGCTCGATT AATTGCTTTG TGTGCGACGA AAAAAACAGA	9120
AATGACATTT ACTCACAGTG ATAAGATTGA CCGAGTGGTA ACCAATCGGT GGTGGCGTT	9180
GCCAATCTTT GCGTTTGTGA TGTGGTTGGT CTATTACTTA TCTATTCAAA CGGTTGGTAC	9240
AATGGGGACA GACTGGTTAA ATGATGTCTT TTTTGGGGAA TGGGTTCTC AATTTGTGGG	9300
AAATTGGTTG GCGCAATGgC AAGTAGCGCC ATGGATGCAA AGTTTAATTT TAGATGGGAT	9360
CATTGCAGGT GTTGGGGCGG TGTTAGGATT TTTACCACAA CTGGCCGTCC TGTTTTTATG	9420
TTTAGGTTTT TAGAAGATTG TGGCTATATG GCCCGGATTG CCTTTGkCAT GGATCGCTTG	9480
TTTCGGAAAT TCGGCCTATC TGGTAAGTCG TTTATTCCAA TGTTAATTGC GACAGGTTGT	9540
GGGGTCCCAG GCGTGATGGC TAGTCGAACG ATTGAAAACG AACGCGATCG GCGAATGACT	9600
ATTATGGTCA CGACTTTTAT GCCATGTTCA GCGAAGTTGC CAATTATCGC TCTGATTGCA	9660
GGTGCATTTT TTCCTAACCA AAGTTGGGTC TCGCCCTCTG CTTATTTCTT AGGGGTTGCT	9720
TCCATTGTGT TATCAGGAAT TGCTTTAAAA AAGACCAAGC ATTTTTCGGG AGATCCAGCA	9780
CCATTCATTA TGGAACCTCC CGCCTATCAT TTGCCACAGC TTAGAAGTGT GGTTCGGCAT	9840
GCCTATGAAC GATGTCGCTC ATTTGTAAAA AAAGCAGGTA CGATTATTTT TGTTTCCAGT	9900
ATTCTCATTT GGTTCATGTC ACATTATTCT GTGACGTTT AACCCTTCC AGAAAGTCAA	9960
AGTATTCTAG CATTTCTAGG TAAAGGGTTG GCGGTGCTAT TTATCCCATT AGGCTGGGGT	10020
AATTGGCAAG GGGCAGTTGC CACAATTACA GGATTGATAG CGAAAGAAAA TATTATTGGT	10080
ACCTTAGGTA TTTTATTTGG CAATGTGAAA GATGTTTCTG AAAACGGGGT AGAAGTCTGG	10140
GGCGCACTGC AACACACCTT TACGCCAGTC GCGGCCTATT CGTTCTTAAC TTTTAACTTA	10200
TTATGTGCAC CTTGTTTTGC TGCAATTGGC GCAATTGCTC GTGAAATGGG AGACCTCAAA	10260
TGGACATTAG GAGCCATTGG TTATCAATGC GGTTTGGCTT ATATGGTCAG TTTTGTGATT	10320
TATCAATTAG GCCATGTGTT AGTCGAAAAA GGCACACTAA CTCTGGGGAC TTTTGTAGCG	10380
ATGGGTGTTG TTTTAGCTGG ATTTTATTTT CTTATAAGAA AACCAGGCC AGACAAAGAA	10440
AGTGTCGAAG CAATCACTTC TTTAGAAAGA GGGTAATAAC GGATGGCAAC AATTATTTTA	10500
TCCATTTTAA TTTTGGGAAC AGCAGGAATC ATCACCTATC GACGAATCAA AAAAGGGGAA	10560
AGTTGTGAAG ATTGTCAAAC AGCTTGTCCT GTAAAAAAG AACAATCTTC ACAaGAATAA	10620
CAACTGAAAA AAATCCCTC TTGTATAAAT AATAGCGGAG TGCTACAATA TTGAGGTAGC	10680
TGAAAAGCCA TTTTATACAA CGGGGGCATA GCTCAGTTGG TTAGAGCATC TGACTCTTAA	10740
CCAAATGCTC TGCTTGTGAA AAAAGCTTTA TAAACTTTA TAAATAGGTT TTTAGTTTGT	10800
GAATTTTTTT ATTGCCAACA TACTGGCAAC AAATTGCCAA CGAAAAAAT TCTAAAAACA	10860
AGAAACGATA CCACGAAGAT ATTGTTTTTT TGTTTTTTAT TCTTTAAACC CAAATGCGTT	10920
TTCTAATGCT TCTGCTGCAC GCTCTTTTAC TCCTGCAGTT AAATGTGTAT ATAAATTAGC	10980

AGTTGTCGAA TACGATGCAT GTCCCAGCCA TTCTTGAATG GATTTCATAT TCTGTCCACT	11040
ATCTAACATT ATTGAAGCAT TAGAATGACG CAAGTCATGA AAGCGAATTT TTTTCAATCC	11100
ATTTTTCTTA ATCAACTTAC TCAAGCGATG TGTTACATAT TGAGGCTTCA TCAACCGTCC	11160
ATCTTCCATA GTAAATACGT ACTCACTATC TTTTAAATA TAATCCTCTC CTAAAAGAAT	11220
ACGATTATCT TGTTGCCATT TAGCTTGATT TACTAATGCT CGTTCAATCG TTTCAACCAA	11280
AGGCAACGTA CGATAGCTGG AAGAGTTCTT TGTTTGATCT TTCTTTATTA ACTGGGACAT	11340
ATTCTCTGGA TTTTCTATGA TTACATGGTT TATCATGATT TTTTGTTTA AAAAATCAAT	11400
AGCAGACCAT TTTAATCCTA AACTTCACT TCTTCGCAAT CCATAAAATG CTGTTAAAAG	11460
AACTACCAA TAGAGCTTTT CTTCTGACAA AATATCAAGT AATTTCAAGA TCTCGTGTTG	11520
ATTATAAACT TGAGAAACAA ATTGTTTCTT TTTTGGCTTT CTTACTTTAT CTGCTACATT	11580
TGTACTTATC AATTCTGTAA TAACAGCCTC TTGAAGAGCT TTGCGTATAA ATGCGTGATA	11640
ATGAATGACC GTATTAGCTG TACAGTCATC ATTTAACATA TATTGGTAAA AATCTTGTA	11700
ATGATAAGGC TTTAACCTAT TTAGTTTAAT TTTTTTAGGA TAGAAAACT TGTAATTCG	11760
ATTATGGACA TAATGCTTAT ACCCAACATA TGTGTAAAT GCTACTTGTT TTTTGTATAC	11820
TTCTAAATAA TTAAGCAAAT AGTCATAAAA CAATAAATCT GGATTAGCAC ATTCACATTC	11880
CATAGCTAAT TGACTTTCCA ATtCTAGTAA AAATTCATCT CTAATTTTTT CtGCTCGTTT	11940
TTTATTTCTT TTCTTTTTTA ATTTTGTATC TTTCCACTTT TGTCTAATTT TTCCGTTAGA	12000
GTCTTTATAA GATACAATTA CTTGTAATGT TCCTCTTTTC TCAATTATGT TTGCTCTTAA	12060
TTTCAATTCA ATCTCTCCTC TTAACTGAA ATATCGGCAA AAATTTTGAC TATTCTATTA	12120
TAGCATGATT TTTACCTTCA TCTTGTATTT TTAATATATA CTCAATGACA GCTTTTTTAG	12180
GTATTTTGTA CCTTGCTCCA ATTTTATAT TTTCTATTGA TTTATCTTTC AGCAAATTAT	12240
AAGCTTTATT TTTTCCAATA TTTAACATTT TAGTTAAATC AGTAATTTTC AATATCGAGG	12300
GATAACTTTC AAATATTTTT TCATAATCTT GAAAATCCAT AAAATTCCTT CCTACATTCT	12360
GATAATAAGT ATTGCACTTT TTTATTTTAG GTGTTAGAAT AACAAAGTAT TAATTTGGGC	12420
TGTTAGCTCA GTTGGTAGAG CAGATGACTC TTAATCATCG GGTCGCGGGT TCGAGCCCCT	12480
CACAGCCCAT TGGGTGCCAA ACCCTTAAAT ACAGTGTAAG CGTTGGTGCT TAGCATTTAG	12540
AAAACGTATA TACTGTATTT TTTTGTTTTA AAAAAGACT GTAGTACCCC GCTTTTTTAC	12600
TAAAAGCCCA CTGTAGTCAT TGATATTTTT GATAAGAAAG GAACATAATT ATGAGTAATG	12660
AAATTGCTAA TATAACAATA ATCCGAGAAA CTTTACAAAA CTATACTGCA AATGAAATCT	12720
CTAAACATAC AGGATTAAAT TTGAGTACTA TAAAAACTA AAAAGTGAG AACGTTTAAT	12780
AGAAAACTA AATTACATG ATGCTATTTG TCTCACTGAA TTTGGTTTTA AAAATAATAG	12840
AAAAAGTGTT GAAATCAACA TATGGAAGTA ATTCTTAATC AGCTGCTTTT GTCAAAAGCA	12900
CATAGGAATT TCACCTCTCT CCAAGTTTAC GGAGAGCCTT ACGGAAGTAT CATTATCCCA	12960

AATTCAAATC ATCGCTTtTT ATgACTATAA AAATtaTTTT AGAATGTTtA TCGCTCCTTC	13020
TTTACGATAA GAAATCGTGG CACAACATAA ATATGCTCCT TGGAGTGTTA ATGTACTGAT	13080
AGAATGTATT TAATTTTCAA TGATCAATAG GTTTTCTACA TAAAAGAATT TACTAATCTT	13140
TTAAATATTA GAAGATTATT TTAAACAGAA AAAGTGGAAG GGGGCTTTTT TTAGTTAGTT	13200
AATTTTAAAT TTAAAGATAG CTTCTATCAA TTTAATTTCT AGCTGTCTTA GCATATACTC	13260
ATCTACATAG GTAATAAATT GCTCCGCCCTC ATTAAATCCA TTTGTCAAAC ATAACCTAAC	13320
TATATAACTA TGATAATGTT TGAGTATCAT CTTGATCGAA ATTGGATCTT CTTCTAGAGC	13380
AGATTGAATA ACGGATAGCG GAACTAATGG ATACTGCCGT TTTTTTTTGT TGTCTACCAT	13440
AGGTCATACT CCGTTTTTAA TTGTTCAACT GCCTTTTTTTC GTTGATACCA GATACCACTT	13500
GTACTCATGT TGAGAATTTT ACCGATTTCA GTATTATTAA AACCTGCAAA ATAATAAAGT	13560
AGGATGACCT TTCTTCTTAT CTCTGATAAA TGATCAATGG CTGTTGATAA CTTTGTATCT	13620
GAAATTAATA TTTCCATTCC AAACAAGAAA TAGACATCTT GTTCATCAAA ATGAGTTTGA	13680
GAGTACTTGC CtTCAGTTGA ATCTGTCAAT GGTTCCCTCAT TATACTTTAT GTTAGkGTTT	13740
CGTTTTTCGAA TAGACGTAAT TTCGTTTCGT AATACTTTTT TGCAGTAGGC ATCAAATTGG	13800
TGTTGTATAG ATTTCTCATT ATAATTTTGT TGCATGATAA TATGCTCCTT TTTCCCCTTT	13860
CCACTAATAC AACAAATCAGA GAGCCCTTTT TTTTCGAAAA AACTAGAAAA AAACAAAAGA	13920
CACGTAGTTT TTATCTACGT GTCTTCCATT TTTTTCAGAA CAACTGCAAA GAAAAGAAAT	13980
ACATATTATA CAAAACAAAA AAATATTCTA ATTTTTTTCA TATAGAAATA ACTCCTATTT	14040
TAAAATTATA AATCGTAACG AATGCGATTT ATCCAATGTA ATTTTACATA CTTTATTTTA	14100
TTTGTCAACA AAAAATGAAC GATCATAATA ACGTTTTCCA AACTTATTTT GCTAGATACT	14160
ATTTTAATTA TTCTTCACTA TGACATCATA ATAAAGTTT AATCAGAAAA TTTAGACGAT	14220
ATGAAACGGA ATAATCAATT AAAAGCTAAA TATGCTTTGA GAAAAAATAA TTTGTTTTTA	14280
GCGGAAGACA TATAGATAAA AACTATGTGT CTTTTATTTT TCTAGCTTTT TCGAAAAAAA	14340
AGGGCTCTCT GATTGTTGTA TTAGTGGAAG GGGGAAAAAG GAGCATATTA TCATGCAACA	14400
AAATTATAAT GAGAAATCTA TACAACACCA ATTTGATGCC TACTGCAAAA AAGTATTACG	14460
AAACGmAATT ACGTCTATTTC GAAAACGAAA CmCTAGAATA AAAGAAAATG AGGAACATT	14520
GAACGTCTCA AATGAAGCTA AGTACCCCTTA CTCATTTTGA TGAACAAGAT ATCTATTTCT	14580
TGTTTGGAAT GGAGATATTA ATTTCAAATA AAAAGTTATC AATAGCCAAA CCAGTTATTG	14640
GACATAAAAA ATTTTTTATT TTGCAAGTTT AATGATACTG AAATCGATAA AATTCTCAAT	14700
ATAAGTATGA ATGGATTTAA TGAAGTGGTC AATTTATTAC CTATGTAGAA GAGTATACAT	14760
TGAGGGAGTT ATAAATTTTC AGAATCTATA TTGAAATTTA AAGTTAAATA ATTAATAAAGT	14820
CCCCTTTCAA TTATTTTGGT AGAAAATTTA TAGaTCATTG AAAATTAAAT ACATTCTAGC	14880
AGTACGTAA CACTTCAAGG AGCATATTTT AGTTACGCCA TGATCTCTTA TCGTAAAGAA	14940

GGAGCGATAA ATATTCTAAA ATAATTTTTA TAGTCAATAA AAAGTGATGA TGTGAAGTTG	15000
GGATAATGAT ACTTCCGTAA GGCTCGCCGT AAAGTTGGGG AGAGGTGAAA TTCCTATGTG	15060
CTTTTGACAA AAGCAGCTGC TTAAGAATTA GTTTTAAGAG TAAATTAAGT AAGACAAAAA	15120
GCATCATGTA AGTTTAGTTT TTAAGCACTT AATTTTGTAT ATTTAGAGAT TTCATTTGTA	15180
GTATATTTTT TAGTTTCTTG aATTTTAtTA TATTAACtAT TTCATTACTC ATAATCATGT	15240
TcTTTTTTTT AAATcTAAaT GATTATAGTT TAGTAAAAAA GAGAGGATGG AGTCCCTTTT	15300
GAGAATATAA AATACGGTAT ATACGTTTTT TATGCTAAGC ACCAATGTTT ACTCTGTATT	15360
TAATAAACTA ACATAAGAGA CTATAACCTT TGACCCGATA ATTATGATTA TAGTTATTAG	15420
TATAAATCTT TAAGTAATAG AAAAGATGTG TTGAATGAAG TATACATTCT GTAATTAACA	15480
ATGAATCATC AGTCATTGGA TAGGAGGCTT TATTTTAAAA GTGTTAAACG ATTCATCTTT	15540
AAGGTGTAGA CAAGAATAAG TAAAAATAAA AAGCTAACTA TCTCTCGAAA TTAAATATTC	15600
TACAGAATGT TGTTTAGAAA AAAGGACTAG TAACAAAAGC TACCTACATT AGGTAGCTTA	15660
AAAGAGTGAA AAGAATATAT GTTAAATTAC TAATTTTATA GCATAAAGTA AAATATTTGT	15720
CAAAAAAAG GGGGAGGGGA GAAGGATGTA TTGTGCACGT GATCCATGCA GAAAGAATCA	15780
TTGCCTATAA AATAmAAAAA TTTACGAGGA GAATGATTCT TGAAGCAACA AATGAAAGCA	15840
AAAAACAAT ATAAGACATA TAAAGCTAAG AATCACTGGG TAACTGTCCC TATTCTTTTT	15900
CTAAGTGTGT TAGGAGCCGT AGGATTAGCT ACTGATAATG TACAAGCCGC GGAATTAGAT	15960
ACGCAACCAG AAACAACGAC GGTTCACCC AATAACCCCG ACCTGCAGTC AGAAAAGGAA	16020
ACACCTAAAA CGGCAGTATC TGAAGAAGCA ACAGTACAAA AAGACACTAC TTCTCAACCG	16080
ACCAAAGTAG AAGAAGTAGC GCCAGAAAAT AAAGGTACTG AACAAAGTTC AGCTACCCCA	16140
AATGATACCA CAAACGCGCA ACAACCAACA GTAGGAGCTG AAAAATCAGC ACAAGAACAA	16200
CCAGTAGTAA GCCCTGAAAC AACCAATGAA CCTCTAGGGC AGCCAACAGA AGTTGCACCA	16260
GCTGAAATG AAGTGAATAA ATCAACGTCC ATTCCTAAAG AATTGAAAC ACCAGACGTT	16320
GATAAAGCAG TTGATGAAGT AAAAAAAGAT CCAAACATTA CCGTTGTTGA AAAACCAGCA	16380
GAAGACTTAG GCAACGTTTC TTCTAAAGAT TTAGCTGCAA AAGAAAAAGA AGTAGACCAA	16440
CTACAAAAAG AACAAGCGAA AAAGATTGCC CAACAAGCAG CTGAATTAAA AGCCAAAAAT	16500
GAAAAAATTG CCAAAGAAAA TGCAGAAATT GCGGCAAAAA ACAAAAGC <sub>r</sub> GA AAAAGAGCGy	16560
TaYGA <sub>r</sub> AAAG AAGTCGCKGA ATACAACAAG CATAAGAACG AAAACAGCTA TGTCAATGAA	16620
GCGATTAGTA AAAACCTAGT GTTCGATCAA TCTGTCGTGA CGAAAGACAC TAAAATTTG	16680
TCGATTAAAG GCGGAAAATT TATCAAAGCA ACTGATTTTA ATAAAGTAAA TGCAGGGGAT	16740
TCAAAAGATA TCTTTACAAA ATTACGGAAA GATATGGGyG GGAAAGyTAC TGGCAACTTC	16800
CAGAATTCCT TTGTAAGA GGCAAATCTT GGGTCTAATG GTGGGTATGC GGTTCTTTTA	16860
GAAAAAATA AACCAAGTAC AGTGACCTAT ACAGGACTAA ACGCTAGTTA TTTAGGACGT	16920

AAAATTACAA AAGCAGAATT TGTTTATGAA CTACAATCCT CACCAAGCCA AAGTGGAACG	16980
TTAAATGCAG TATTTTCAAA CGATCCGATT ATCACrGCTT TTATTGGTAC AAACAGAGTC	17040
AATGGTAAGG ATGTTAAAAC ACGCTTAACG ATTAAGTTCT TTGATGCGTC AGGTAAAGAA	17100
GTACTACCAG ATAAAGATAG TCCATTTCG TATGCGCTGT CTTCTTTAAA TTCAAGTTTA	17160
ACGAATAAAG GTGGCCATGC GGAATTTGTT TCTGATTTTG GGGCsAACAA TGCGTTCAAA	17220
TACATTAATG GytCrTATGT GAAAAACAA GCGGATGGAA AATTTTACTC ACCGGAAGAT	17280
ATTGACTATG GCACAGGACC TTCTGGATTG AAAAATAGTG ATTGGGACGC TGTAGGTCAC	17340
AAGAATGCCT ACTTTGGTTC AGGTGTAGGT CTAGCyAATG GrCGTATTTT CTTTCTTTT	17400
GGTATGACAA CAAAAGGAAA AAGTAATGTG CCTGTATCTA GTGCGCAATG GTTGCCTTT	17460
AGyACTAACT TAAATGCGCA ATCAGTGAAG CCTATTTTCA ATTATGGGAA TCCAAAAGAA	17520
CCAGAAAAAG CAACGATTGA ATTCAATsGA TACAAAGCCA ATGTCGTTCC TGTwCTTGTG	17580
CCrAatAAAG AAGTCACTGA TGGyCAGAAA AATrTCAATG ATTTAAATGT GAArCGTGGC	17640
GATTCTTTAC AATACATTGT GACAGGGGAT ACGACAGAAC TTGCCAAAGT AGATCCAAAA	17700
ACAGTAACmA AACAAGGGAT TCGAGATACm TTTGATGCAG AAAAAGTGAC GATTGATTTA	17760
TCCAAAGTGA AAGTTTATCA AGCAGACGCA AGTCTrAACG ArAAAGACTy AAAAGCTGTT	17820
GCTGCAGCrA TTAATTCAGG AArAGCTAAA GACGTGACTG CTTCTTATGA yCTcAaTTTA	17880
GATCAAAACA CCGTCACAGC AATGATGAAA ACCAACGCrG ACGGyTCyGT TGTTTTAGCA	17940
ATGGGGTATA AATATTTACT TGTCTTGCCG TTTGTAGTGA AAAATGTAGA AGGCGATTTT	18000
GAAAATACAG CTGTTCACT GACAAyGAT GGnGAAACGG TAACAAATAC AGTGATTAAC	18060
CATGTGCCaG GTaGTAATCC TTCCAAAGAT GTAAAAGCAG ATAAAAACGG TACAGTTGGC	18120
AGTGTTTCTC TACATGATAA AGATATTCCG TTACAAACAA AAATTTATTA TGAAGTGAAA	18180
TCTTCCGAAC GTCCAGCyAA CTATGGCGGA ATyACmGAAG AATGGGGCAT GAATGATGTC	18240
TTGGACACGA CCCATGATCG TTTCACAGGk AAATGGCAGC CTATTACrAA yTATGACCTT	18300
AAAGTAGGGG AyAAAACGTT AAAAGCAGGA ACAGATATTT CTGCCTACAT TCTTTTAGAA	18360
AACAAAGACA ATAAAGACTT GACGTTTACr ATGAATCAAG CATTATTGGC mGCsTTAAAT	18420
GAAGGAAGCA ATAAAGTAGG CAAACAAGCT TGGTCTGTGT ATCTGGAAGT CGAACGGATy	18480
AAAACAGGTG ACGTAGAAAA CACGCAAACA GAAACTACA ACAAAGAGCT TGtkCGTTCT	18540
AATACsGTGG TGACGCATAC rCCTGATGAT CCAAACCAA CCAAAGCCGT TCATAACAAG	18600
AAAGGGGAAG AyATTAAyCA TGGAAAAGTk GCTCGTGGTG ATGTTCTTTC TTATGAAATG	18660
ACnTGGGACT TAAAAGGGTA CGATAAAGAC TTTGCCTTTG ATACAGTCGA TCTTGCGACA	18720
GGCGTTTCTT TCTTCGATGA TTACGATGAA ACGAArGTGA CACCAATCAA AGACTTACTT	18780
CGTGTCAAAG ATTCTAAAGG GGnAGACATT ACGAACCAGT TCACGATCTC wTGGGACGAT	18840
GCCAAAGGCA CGGTGACrAT yTCTGCCAAA GACCCACAAG CCTTTATTCT AGCGyATGGT	18900

GGGCAAGAAT	TGCGTGTAAC	rCTCCCTACA	AAAGTCAAAG	CcgATGTTTC	TGGkGATGTT	18960
TATAATTCAG	CGGAACAAAA	TACATTGGy	CAACGAATTA	AAACCAATAC	yGTTGTCAAC	19020
CATATTCCAA	AAGTGAAyCC	TAAAAAAGAC	GTGGTTATTA	AAGTyGGTGA	CAAACAAAGt	19080
CAAAATGGyG	CCACAATCAA	ATTAGGGGAG	AArTTCTTCT	ATGAATTTAC	AAGTAGTGAC	19140
ATTCCTGCAG	AATACGCTGG	wGTTGTGGAA	GAATGGTCGA	TTAGCGATAA	ACTAGACGTC	19200
AAACATGACA	AATTTAGTGG	cCAATGGTCT	GTGTTTGCCA	ATTCTAATT	TGTTTTAGCA	19260
GACGGAACCA	AAGTGAATAA	AGGGGACGAC	ATTTCGAAAC	TATTCACGAT	GACCTTTGAa	19320
CAAgGGGTAG	TGAAAATCAC	GGCCAGTCAA	GCCTTTTTTrG	ATGCGATGAA	TCTAAAAGAA	19380
AACAAAAACG	TTGCACACTC	ATGGAAAGCG	TTCATTGGTG	TAGAACGAAT	TGCGGCAGGA	19440
GACGTTTACA	ACACAATCGA	AGAATCTTTC	AACAATGAGA	AGATTA AAC	kaATACGGTA	19500
GTGACrCATA	CGCCAGAAAA	ACCACAAACr	CCACCAGAAA	AAACAGTGAT	TGTACCACCA	19560
ACACCAAAAA	CACCGCAAGC	ACCAGTAGAG	CCATTAGTGG	TAGAAAAGGC	AAGTGTrGtG	19620
CCAGAATTGC	CGCAAACAGG	CGAAAAACAA	AATGTCTTAT	TAACGGTAGC	TGGTAGTTTA	19680
GCCGCAATGC	TTGGCTTAGC	AGGCTTAGGC	TTTAAACGTA	GAAAAGAAAC	AAAATAATTG	19740
AAGACGAAAA	GAAGGACGAA	yTGmTTGTCC	TTCTTTTTTT	AGTTATGGAG	GGGAAAmATT	19800
GGAAGCAGTA	GTAgtAGAAA	GAGAGGCAAA	GGGGATGAAA	GAAATTGCCA	TCCAAGAAAA	19860
AGATTTGACC	TTACAGTGGA	GAGGAAACAC	AGGTAAGTTA	GTTAAAGTTC	GATTrAArAA	19920
TACACGTGCA	ATGGAAATGT	GGTACAACAA	ACAAATTACC	GAAGAAAACA	TTCAAGAGAT	19980
CACCACGTTG	AATATyATTA	AAAAYGGAAA	ATCTyTGGCA	TTAGAAGTAT	ATCCAGAAAA	20040
AAGTATCTAT	GTGAAACCAA	ATTTAGGCAG	AATCAATGTG	CCTGTCTTTT	TTATCAAAAC	20100
ACCTATTAAT	AGAGGGATTT	TTGAAGAGAT	CTTCGGCGAA	ACTTTAAAT	CATAAATAAA	20160
AAATTATTTG	GAGGAAATTA	CAATGAAAAA	AATTATTTTA	TCAAGCTTGT	TTAGTGCAGT	20220
ACTAGTATTC	GGTGGCGGAA	GTATAACAGC	ATTCGCTGAC	GATTTAGGAC	CAACAGATCC	20280
AGCAACTCCA	CCAATTACCG	AACCAACTGA	TTCTAGTGAA	CCTACGAATC	CTACTGAGCC	20340
GGTGGATCCT	GCAGAACCGC	CAGTAATACC	AACTGATCCA	ACAGAACCAA	GCAAGCCAAC	20400
CGAGCCTACA	ACACCGAGTG	AGCCAGAAAA	GCCAACAGAA	CCAACAACGC	CAATTGATCC	20460
TGGAACGCCG	GTTGAACCGA	CTGAACCAAG	CGAGCCAACA	GAACCTAGTC	AACCAACCGA	20520
GCCTACAACA	CCAAGCGAAC	CAGAAAAACC	TGTTACTCCA	GAACAACCGA	AAGAACCAAC	20580
TCAACCAGTG	ATTCCAGAAA	AACCAGCAGA	ACCAGAAACA	CCAAAACTC	CTGAACAGCC	20640
CACTAAACCA	ATAGACGTAG	TCGTTACACC	TAGTGGAGAA	ATTGATAAAA	CGAATCAATC	20700
GGCAGGAACA	CAACCAAGTA	TTCCTATTGA	AACAAGCAAC	TTAGCGGAGG	TAACACATGT	20760
ACCAAGTGAA	ACTACTCCAA	TTACAACAGA	AGCTGGGGAA	GAAATTGTAG	CAGTAGATAA	20820
AGGTGTTCG	TTAACCAAAA	CACCAGAAGG	ATTAAACCA	ATTAGCAGCT	CGTATAAGGT	20880



TTTACCTAGC	GGAAACGTTG	AGGTAAAAGC	AAGTGATGGA	AAAATGAAAG	TATTGCCACA	20940
TACAGGAGAG	AAATTCACAC	TCCTTTTCTC	TGTATTGGGA	AGCTTCTTTG	TATTAATTTT	21000
AGGATTCTTT	TTCTTTAAAA	AGAATAAGAA	AAAAGCTTAA	TCAGAAAGAG	TAATTTTCGAT	21060
AAATTTCTGT	GAGGAAAGGA	GGTGCTTTGC	TTGTTTATTT	TATTTACTAC	mAAGCnAAGT	21120
CATGTATTAC	TAAATGAATT	AGGGTTTAAT	AGAAAAATTG	ACCTGGTACA	AGGTATTGAA	21180
CAGTTAACAA	AAATCAATGC	GGATTTAAAA	AAATCGACGG	ATCGAATTCT	AATTACGTAT	21240
GAGATTTATT	ACAACGTGAC	AGACGAGAAA	GCTTTTTTAA	GAGAAACAAT	GACTTTGCCG	21300
GTAGTTGAGG	AAACTGTCCT	CCAAAACATG	GTACAGACTG	TAGAAACCAA	TGAAGAAGTA	21360
AGACAAGAAG	ATTTAGAAGC	GTTTCTTTAC	CTGTTCAAAA	GAAATGTGAG	TAATAAGTCG	21420
AAGAGAATAG	GCCGTTC AAC	AGGTTCCTTA	AAAAAGAAAG	CAGAAAAGCA	ACAAAAATCA	21480
ACTGTGTTTA	CAACAATTTT	TCGCTATAGC	TATTTCTCAG	TGGCTTTACT	TATATTTTTA	21540
GCAGTAGGTG	TTTTTACAGG	TGCGTCATTT	GGTTCTCAAA	GTAAACCAAC	GAGCCAAAAT	21600
GATTCTAAGT	TAGAGGAACG	ATTAACACAT	TTACAAGAAC	AAGTGCATGA	GCAACCACAA	21660
GTTGATGTTG	TTGCTCGGTT	TTTCTTATCC	aGTCTGTACT	CTGGTGAAGA	GAGTGACAAA	21720
GCCACGCAAA	AGCAAGTTAA	AAAGTATGTT	TCTGAAGGTG	TACTCAAAGA	GATTAAGGGG	21780
TCTAAAGAAC	AAATAAGAAC	GTTATTTCTT	TGGGAATCAA	AAAAAGAAGG	CAACCAATGG	21840
CTCCTTACGT	ACGTTGTTAC	CTTGAAAGTG	GCTAAAGAAG	AAACTGAAAT	TCGTCAAATG	21900
AAAGTAGCTA	TGGAAGAGAA	AAACAATAAT	TATCAAGTAA	CAGAACTTCC	AGAAGAAAAA	21960
GAATTTACTA	TTAATGAATA	ATAATTGAGG	AGGCAATTCT	ATGAAATATG	AACGTCCATT	22020
AAAAAGAGAG	TCACAAATCA	AGGAGTTTGA	GTTAGGCACG	CACGCGGCAG	TAATTGAAAA	22080
AGTTCAAAAAG	AAACGGTCTC	AAAAAGGGAA	TGATATGTTT	TTACTCTCTC	TTTTGGGAAA	22140
AAGTAACGAA	AAAGGGGTTT	ATTTTCTTAC	cTTCGGGAAT	GATTATACTG	AAGATAATTT	22200
ACGCTATATT	TTAGCCAGTA	TCCAAGACAA	CGGTGTAGAG	ATTCCAGACG	TTGACTTTGG	22260
CTATAATCGA	GAAACTTTTG	AATTTCTGAA	AGGTAAAGAT	GTCTATATTC	AGGTTGAAGA	22320
mCAAGAGTAT	AAAGGGAAAG	TTAAACaTGC	GGTAACaAAT	TTTTTAACTC	aGGATGAATT	22380
TGAAGAAAGC	GAAGAAATGG	AGTTTTTCAGA	AAGCAATACT	GAAGAAGATT	GGTAAAAATA	22440
GCCATTTCTT	AGAATATAAA	AACTGAAGGA	GAGACAAAAC	TATGGATATA	GTCACAAAAA	22500
TTATTACACT	AATTGGAGGA	ATCATTGGAC	TTGTTTCAGC	AGTTAGTATT	ATGTTTGCGG	22560
TGAAGAAAT	TCGTTCAGGT	ATGTCGAATG	ATGATCCACG	TACGCTTGAT	AAAGGAATAG	22620
AAAAAGTAGT	TGTTGGTGGG	GCTGTTATTT	TAGCAATTGG	TGGGGTAGTC	GCCTATGTTA	22680
TTACACAGGT	AGGaGCTATT	CGTTTTTAAA	AGGAGGAAAA	AAATGAAGCT	ATTAGTAGAA	22740
ATGATAGTCA	ACGGTCAAAC	TGAATGGGAA	GTTGTAGAAG	AAGAAAATGC	GCCACAGGCA	22800
ATTATTCAGA	GTCGAGGAGG	CTTTTCATTT	GATGAGAATG	GAGAGCTCAT	TGTCAATGAT	22860

GATGAAATAA GTTATACAGG CGTGTTTGAG GTTTGCGAGA CTAACCTATT GGATTTCACT 22920  
 GTTAAGGAAG CAGAAATCCA TAGATTTTAT CACAAAAAAT TAGAAAAGCT AGGGAATAGAT 22980  
 CCTCTAACGT TTGAAAATC GCmAGAAATA GCTAACTGAG CGGAGGGGaa GCnATGAGAA 23040  
 ATCCAAAAGG AACGACCGAT CAGCTATTTA AAAGTTTAGC TGAATACmAT CCmACmACgA 23100  
 aTGrAGCCAT GTCTAAAAtA GcGAAAGTCC tAGTCCCTtTT AGGAATCgCT ATTTTAGGCA 23160  
 TTCTTTTTAT GATTGAATTG GCGAATACCC AAAAGAAATT TCAATCAGAA GACGGTGGTT 23220  
 TGACAATCGA GCTATTAAC AACATTGCAT TGAAGTATGT AATCGCGTAC GTCTGTATTA 23280  
 TGGGTCTCG CTACATTATT GATGGCATTG TTTGGTTTAC AATCAAGCA GCCAAATGGA 23340  
 TTAATTCAT TGTAACAGCG ACAGGAACAA GTGAAGCGAT CCCACAAATG GAAAAAGTTT 23400  
 CTTGGTGGGC AAAACCAATT GTATTTTCT TCAAATATT TGCATATATT GCATTAATAT 23460  
 TATCTAGTAT TATTGCAAT ATTCTTATT TCTTACGAGG GATACAATTG TATATGATTA 23520  
 AAGCTCTTGC ACCGTTGTTT GTGGCGTTTT TTGTTTCATGA TGAATTACGT TCGATTGCCA 23580  
 TGGGGTACAT GAAACAAATC ATGGCCTATG CGCTTCAAGG CGTCTGTG GTTCTATTAC 23640  
 TTGGGTAAAT TCCGATATTA ACAGCAAACG ATTATCTATC CTTGAATCT CTTGAAGGTG 23700  
 GCATTTTGgC AGGTGCTGGA GCGATCATCA TTAATGTTAT GACGTACTTT AATTAAATTT 23760  
 TAAAATATGT GGCAGTGATT ATTCTTCTAG TTGGTTCACA AGGATTGACC AAACGATTAG 23820  
 TAGGAGCGAT GTAAAATGGC GATTAGTTCA GAATTCTATA rAGATTGTC AAAAGTAGAG 23880  
 AAAAAATCT GGGGAATTAC AGTCAGAGAA TTTAAAGCGT ATGTTTGCTT TGTTTTTATA 23940  
 GGGATATTC TTCTGTTAGA AGTATTTTTT TTGCCTGATC TTTATTTAT GGTTTGTCG 24000  
 TTAG 24004

## (2) INFORMATION FOR SEQ ID NO: 73:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2707 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

ATTTGAAACC TGTTGTTGAA TCCATGAATA CTAAGGTTG GTAATTTGGA nnGATATCTT 60  
 GTTTCATACG TTTTTCGCT CCTTTGCCCT GATCCATTG CTGTATCAGA GTTATTTTGT 120  
 CGGTATACAA CAATGCTATC TTACCATGTT TTCAGTTGGT ATGCAATTAT CTTTTTGGAT 180  
 TTCTTTTAAAT TTGTTTGCCA AACGAAACAT CATGGAAAAC TTCAAATAAT TGTTGATTAT 240  
 TTTGGGTCCG TTTCAACATT TTTATGAATT GATCCGTATA TTCCAGCGAG TCTCCTGACA 300  
 TATTATTTCT CAATTTCAC GTTCTTCTA GTTGTCTGG TGTCATTAGC AACTCTTCTT 360  
 TTCGCGTACT TGATTTTTTA ATATCAATTG CTGGGAAAT ACGGCGTTCT GCTAATTCAC 420

GAGATAAATG GAGCTCCATG TTGCCTGTAC CTTTGAATTC TTCTATAGATG ACATCATCCA	480
TTGGGCTACC TGTATCCACC AATGCCGTGG CTAAAATAGT CAGACTGCCC CCTTCTTCGA	540
TATTTCTGGc AGCACCAAAA AATCGTTTAG GTTTGAAAAA CGCAGCTGGa TCTATTCCGC	600
CACTTAAGGT TCGACCACTA GGCGGGACCA CTAAATTATA GGCCCTTGCT AAACGAGTGA	660
TGCTGTCCAT TAAAAATACT ACGTCCCGTT TGTCTTCAAC TAAGCGCATT GCTCGATCCA	720
GAACCAATTC AGCAACTCGC GTATGATTTT GTGGTTGTTG ATCAAAAGTT GAAGAAACCA	780
CATCGCCTTT AACGCTTCGC TCTAAATCAG TCACTTCTTC AGGGCGCTCG TCAATTAACA	840
ACAGAATCAA TTCCACATCA GGATGATTTT CTGTAATCCC ATTAGCAATT TCCTTTAAAA	900
CACTTGTTTT TCCCGCTTTT GGCGGCGCTA CAATTAAACC ACGTTGCCCA AAACCAATTG	960
GGGCAAATAC ATCAATCATT CGTGTGATA ATCTGCCAGC CGTTGTTTCT AACGTTAATT	1020
GTTTTTCAGG ATACAACGGT GTTAGCGCTG GAAAAATGTTG CCGCTCTTTC GCTTCTTCAG	1080
GATCTTTCCC ATTGACACTT TCAACATGCA TGAGTCCATA GTAACGTTCT GATTCCTTG	1140
GCGGACGCGC CTTGCCAGCT ACTTTGTCCC CATTTCTCAA TCCGAAACGA CGAATTTGAq	1200
AAGAAGAAAT ATAAAtGTCT TCTGCACTTG GTCCATAATT AATTGGTCGC AAAAAGCCAT	1260
AACCATCCTG AGAAACAATA TCTAAAATTC CTTCCATGAA GAAAAACCCT TGCTTCTCAG	1320
CTTGTGCACG AATCACAGCT AAAGACAATT CTTTTTGGT CATTTGGCAT AATAAGGATT	1380
TTAAcTCTTt CGCGTAqCAT AAATATCTTt TAACGTACTG tTTTCCaATt CAGCCATgTa	1440
AATAGTCGCT CATTGATCCA CCTCAAnCGG ACCTTCTTGT TCTTCGGTTT CTAGCATTTG	1500
AATATCAGCG CCTAAAGCAG TTAACTTTTC GATAATGTGG TCGTACCCAC GTAAAATATA	1560
TTCTACATTA TAAATAGTGG TCGTACCTTC AGCCATCAAG CCGGCAGTGA CTAGGCAAGC	1620
GCCTGCCCGT AAATCAGAGG CGACTACTTC TGCACCGTGC AATTGATTG GGCCATTTAA	1680
GATAATCATA TTGCCTTCAA CCGAAGCATC AGCCCCATC CGCACCAACT CAGGGATATG	1740
CTTATTTCTG TGAGCATAAA TCGTATCAAT AATTTACCA GTTCCTTG TGCTTTAAAG	1800
TAATGGCGTT AACGGTTGTT GCAAGTCTGT GGCAAAACCA GGATACGGAT AGGTTTTCAC	1860
CGTCGTCATT TTCAGATCAT GTGATGGATG AACTTCAATC ATGTCTTCTT CAATTGTCAT	1920
CTTAACGCCC ATTTCTTGCA ATTTGCAAT GAAGCTTTCT AAATGCTCAT AAATAACATT	1980
TCGAACCTTG ATTCCTTCAC CCATAGCTGC TGCCATAGCT AAATAAGTCC CAGCTTCAAT	2040
TCGATCTGGA ATAATCGAGT GACGACAGCC ATGTAATTCT TCGACTCCTT CAATCCGAAT	2100
AACATCCGTT CCAGCTCCTC GAACTTTTGC ACCCATGTTA TTTAATAATG TAGCAATATC	2160
AATAATTTCT GGcTCACGCG CAGCATTTTC AATAATCGTT TTGCCTTTGG CTTTAACTGC	2220
GGCTAACATG ACATTGATTG TAGCGCCAAT AGATACCATA TCCATAAAAA TTCGGGTTCC	2280
TTGTAAGCCA GCTTCGTCTG TTCTCAAGTA CATCGCCCCA TGTTCTGTTG TTACATGAGC	2340
GCCTAATGCT TCAAATCCTT TGATATGTAA ATCAATTGGA CGAGGGCCTA AGTAACAACC	2400

GCCAGGTAAG CCGACAACGC CTTCCCCAAA TCGACCAAGA AGTGAGCCCA TAAAAATAATA	2460
GGATGCACGT AAGCTATTGA TTTTTCCTTT TGGCATTGGC ACAGAAACAA CTTCTGTTGG	2520
GTCAATAATT AATGTATTAT TTTCAAAAGT AATTTTAGCA CCCATAATTT CTAAAATTTT	2580
AATTAATGAG TGAACATCCT GAATGTCAGG AACGCCATCT AAAGTCACGG GCGAATCTGC	2640
TAAATAGCA GCTGGAATTA AAGCGACCAC ACTATTTTTG GCGCCGCTAA TGGTTACTTC	2700
GCCTTTT	2707

## (2) INFORMATION FOR SEQ ID NO: 74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10636 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TCACCTTATT GATAGCGCTT CCTTATATTG nAATAATGCG ATTTTCTAAA CAAAAAATTA	60
TAGGTGGTAT TGATAATGAA AAATCCGTAT GTCCAAATAA CCGACCAAGA ACAATATAAA	120
AAGTCAATTC GTCGAATTCC TCCAAAAGAT CCTAATATTT TTTTAATTGA GAGACCTACA	180
GCAGGTGAGA TTATTGCTCT ATGGTGGACA ACCTCTTCCC GTGGCAAATC AACTGATAAA	240
TATATACCAA AATATTTCAA GCGTATCTAT CGAGTAGATA TGATTAAACA ACGTAAACTA	300
TTCATCAAAG CTAACGTTAT CTCCGAAGAA AATAATACTT ATGTACTGAC TGAAC TAGGA	360
GAAATGTTAG TACATAAATA CTTTATATT ATTGAAGAGC ATAATAATTC TTCCAATGAT	420
TCTTTCAATA CTGAACGAAG AAAGTTAGCT GATAAAATTC TATATTCGCT AGGCGGTGAA	480
ATATAATGCA AACGGAAGTT AAAGAGTTTA TTGCTTTTGA TATTGAAACA ACAGGTTTAG	540
ACTTCCATCA GGATGCTGTT ATTCAGATAT CAGCTGCAAG ATACGTTGAT- GGTATTGAAG	600
TTGATTATTT TGATACATTG GTAAATAGTG ACTACATTCC AGATGAAATT ACCAAATTAA	660
CAGGAATCAC TAATGACCAA GTACTAAATG CGCCAACTTT AGATGAAGTT ATGCCTTATC	720
TTTTTGATTA TTTAGGAGAT ACGATTCTAG TTGGACATAA TATAAAAAGT TTTGACTTTC	780
CTTTTTTAAA AGCAAAAGGA TATAATATTG CAGAAGGACA TGAAATTTAT GATACTCGCT	840
ATTTTGCAGC CACACGAAAG CATGGGGCTG TAAACAATCA ATTAAC TACA TTGAAATTAC	900
TTTTTGATAT CGATGCGATT TCACATAATG CATTGAATGA TGTTAGGATA TCTGCAATTG	960
TATTTATGGA ATTATTAAAG ATAGAACCTC AAGAAAAAG CTCTACTCTT AAGAAAAGTA	1020
GCTTGGACTC TTTTGAATTA GATGATGATA CTACACCTTT TTTTGAAGGA ATGTCCTTTG	1080
TAGTAACGGG GGCATTTGAT GATTCTAAAT ATAACCGTAA ACAAATTGAA AACTGATAA	1140
AACAACATGG CGGTAGGGTA TCGTCAAGTT TATCTGCCAA GACTGATTAC TTTATCCAAG	1200
GTATTCAAAT TTCTTCCCAA CTAAAGACG GAAAGCACAG CAGTAAAGAA CTAAAATATA	1260
TAGAGTTACG TGAAAAAGGG GTTGATATTT ATAAATTTAA TGTAATCAG TTTTATGAAC	1320

TAATTACAAA CTATAGAAAA TTGAGGAAAA ATTAATGACA GAACTAAAGA AAATAACTTC	1380
AAAAAATGGA CATGTTTACC TCTTTAAAGA TATACAACAT AAACAACGAA ATGAGTTAGG	1440
CATAGCCTCT GGACAATCAG TACTGATTTC AATTTTTGAA CCGGATCCAT TCTATTTTAC	1500
AGAAGATATT CAACTTATTC ATTATAAAGA TATCGTTAAA ACCGAATCCG TTTCTTCAAA	1560
AGAAAAATGAA TTTTTTGAAT ATGTAAAACA AAATTCACCA GTTGCTTATA AAAAGAAAAT	1620
CTCGGATAGT TATAGTATCT CTGAGTATAT TAATTATGCT CCACAACCTA AATTTAAGTA	1680
TATTATAAAA GATAGTATAC TAACTATTAC GGGGGAAATT AAAGACAGAA ATGTAACTT	1740
TCCAATTTCT CAGCCTTACG TATATGTCGA TAAACCAGAA AATGAAATTC CGTTAGATAA	1800
TAATTTAAAA TTCCAATATG TAGTCAACGA TTTCGATATA CATGCTCCCG AATACAAAAT	1860
ATCGTTTATT TTAGGTACTT CAAAAGCCGA GTATGGTCGT TCTTTCAATA TGAAAAATTT	1920
CGAAGGTGTT TTACGCACCC CATCTGAAAT TAGTTCTATG AAAGCAAAA AAAAGAAATA	1980
GTTACTTAAA TAAAGCAAAA AAGTTCACG CCTAAAATTC GAATTTAGAC GTAGAACTTA	2040
TTTATTTCAA TGTCCGAACC AAATAAATCT CTTACAAAA GCCTTTAAGT CCATTATAAA	2100
TCCGTTGTGC TCGTGTTTTC TTTTCACATA AAGCAAAGAC TGTCGGTCCA CTGCCGCTCA	2160
TCAACGCTGC ATCAGCGCCG TACTTCAGCA TGCATCCTT GATTTGTTGC ACGATTGGAT	2220
GTCTGGCAAT CGTGACACTT TCCAACGCAT TGCCAACCGT TTGGGTCATC CCAGTATAAT	2280
CCCCATTTTC AATAGCTATT CTTAGTCCAG CAATATCTGG ATGGTGCAAT TCATCTACCG	2340
CTAAATCATT AAAAACGGTA CTAGTGGATA CACTAATCTT TGGTTTAACT AACACAATCC	2400
AACATTGAGG CATGGTCGGT AATGCTTCAA TTTTCTCCCC ACGTCCGTTT GCAAAGGCTG	2460
TTCCGCCCCCT TAAACAATAA GGAACATCCA TACCAATTTG GCTGCCAATT TCACAGAGTT	2520
CGTCCATCGT CAACCCGAGA TTCCACAGTT TGTTCAAGCC TCGTAATGCT GCGGCACAGT	2580
CACTGCTCCC CCCTGCTAAA CCAGCTGCCA CGGGGATACG TTTTCAATA TATATTTTAA	2640
TTCTTTTATG AATATTATAG GTACGCTTTA ATAAGTCAAC TGCTTGATAG ACATGATTCC	2700
GTCGATCCAC GGGTAAAAAA GAGCTGTCCG TTTCAATACG AATGATGTCT TCTTCTAGAT	2760
TTTCAAAAGT TAAATGATCG GCTAAATCTA CACTTGCAAA AATAGATTCA ACTTCATGAT	2820
AGCCATCCAC ACGTTTATGC AGAACATCTA ACCCAAGATT AATTTTCGCA GGAGCTCGTT	2880
CGATTATTTT CATCGGTTCC CTGATTACCC TTCCTTAATC TTTTGTCTAC CAACATTTTA	2940
ATCTTTTAGA TTCTATCATA TGTCCATGAA GAAGTACACC GAAGATTGCT CTTTCTCTGA	3000
CGACAGATGG CTGAAATTAA AAAGAGCTTA GGAACGAAGG GCTCTTCCTT CACTCCTAAG	3060
CGGACTTTCT AAATTATACG GCTTGACCGA TGAATTCGAT TTCAACGGTG CGTGTAAAAA	3120
CATCTGAATA GCTATAAGAT ACACGTTCAA ATGAGTTCTC GTCTGGATCC AAGTCAACAA	3180
CAAAAACAGA TGGATAAGTC TCCGTTAAGA TTCCTTTACG TTCTGTTTGG CGCTTTCTTC	3240
CAGTCTGAGC AACTAACATA ATTTCACTGC CGATACGGCC TTCCAAATCT TTCTTAATTG	3300

ACGATAACGT TGTGGCATT GCATTACCT CGAGAAACAT TATATCACTT TTTTAAAAA	3360
ATTTCAAGTT TATCATACTA GAAGTCTGTT TGCAACTACT TTTGCTCGAC AAACCGGCTG	3420
TTTTGTGAA TTTTATTCA TTTTCTAATG TATATTTTCA TTTTGCTTTG GCACGATCAT	3480
AGGCACTTCG GACAGCTGAA AAAGTTCCCC CACGTACTTT AGCAATTTCT TCTAACGTGT	3540
AGCCTTTTAA GTATGCGTT AAAATTTGTT GTTCCAACGA GGAAAGACGT GTCGGAAAAAG	3600
CTGCCAATTT TTCTTGAATC ATTAATTGAC CTAATGGGTC GGTAGCCGCT GTCCCTAAAT	3660
AGTCCACTAA CTCTTCACCA GCGCTTTCTA ATCGTTGTTC CAGAGAAATA GACAAGGTAT	3720
CTGTTGCGCG CTTATAGGCG CATTGATGGC GAATTAAGCT ATGTACGTGA TTTTCAAAAT	3780
TTCGTTTAAA AAACCTTCCT AATGATGCTT GTTTTTCTGA ATCAAATCGT TCTAACGTCC	3840
GAAAGAAGAC AATTCGCCCT TCTTGAAGCC AGTCCTCAGT ATCATAATCT TTTAAATAGT	3900
ATTTTTTCAT GCACTTATAA ATCACTGGCT GATATTGCCC ATATAAGGCT TGAAATTGTT	3960
CTTGGTCTCC TGCTAATGCT TGTGAATCT GTTCTAACAT AAAAAGCCCT CCTTTTCGAT	4020
AAACGTATCG AATGATTTAG GAGAGCATAG CATGTTAGAA ATTTATGACT TAAGCGAACA	4080
TTTGTTTAAAT CTTTTTTTTG AGACAATTCT TCTAGCTTCT CAGAAAGTTT TGATAACTGC	4140
TCCACGTTCC ATGGAGAATT TCGACGAAAT TGCTGGAAAT GGATATCCGT TGCATGAATG	4200
GCAATCGTCT TTTCTGTTTT CTGGACAGCT TTGTAAAGTT CATTGGCCGA AGTCCGCAAT	4260
GCACCTTGTTG AAAAGACCAC CCATTGTTCA GCCAAATCAC TCGTTGCGAC TGTTACTTGC	4320
GTAAACGAT CATTTTTTTC ACCAGCTAGA CGTTCAATAT AACTATCTGC TGTTTCATCT	4380
TCTTTGGTGA AAACCACGGT CAAAAGGTAC TTTTGTACG TTTGCTGAAT ACCTGGGACT	4440
AATTGGGCAT CAAAGACCAC AATGATTTC AACCCTTCAT ATTTGGCGTA ATTGGACAGA	4500
CGACTAAGAA GCAGTTCACG CGCATCCTCT AAACGCTCTT GCTTTTTCAA CTGGACCAAT	4560
TCTGGCCAAG CACCAATCAT ATTATAGCCA TCAACGATTA ACAATTGTTT TTTCATACAA	4620
TCGCTTCTT CTATAATGGG TTGCGGCCAC GATAGACTTC ATACATTAGC AAGCCTGCAG	4680
CCACACCAGC GTTTAAGCTT TGAACGTGTC CAGTCATTGG AATCGTTAAT AATTCATCCA	4740
CTTCTTTATG AAGACCTTGA CTCATGCAC GACCTTCATT TCCGATGATT AAGGCAATTG	4800
CGCCTTGTTG ATTCCAGCGA CGATAATCCG TCCCAGACAT ATCTGTACCA AAAATCCAAA	4860
ATTGATTCTC TTTTAGCGTT GCAATACTTT GCGCTAGATT CGTTACACGA GCCACTGGAA	4920
TATGTTCCAC GGCCCTGTT GACGCTTTGG TTACGGTTGG TGTAATTCCC ACTGCACGAT	4980
GTTTCGGAAT GATAATACCG TCCACGCCTG TCGCATCAGC TGTTGCAAA ATTGAGCCAA	5040
AGTTATGTGG GTCCTCTAGA CTGTCTAAAA TTAGGAAAAA TGGTGTGCT GTCTTTTCTT	5100
TTGTTTGTTT CAATAATTCT TCTAAAGTTA AATATTGATA CGCCGTGATT GCTAAAACCA	5160
TGCCTTGATG CACGCCGTGA TCACTCATCG TATCTAATTT TGCTTTAGGC ACCCATTTAA	5220
CTGGAACAGC CTGTTCTTTC GCTGCTTGTT TTAATTGCTC AATTTTCTCG CCACGAGCAT	5280

CTTCTTGTA	AAACAATTTA	TTCCCGCGTC	CTTGCTGCAA	CGCCTCAACA	GTGGCATGGA	5340
AACCAAAGAC	AAAGTTATCT	GCTAGCTCTT	CTGAGGTTCC	ATCAGCACCT	GCTGTGCTG	5400
GTCGTGGGTT	CCGTTCTTTT	GAACGCTTAT	CTGAACGTTT	CTTACCTTTT	TTAAAAGGAC	5460
GAGATTTTTC	ATTTTTCATT	GGTTTCACCT	ACTTTCTTGA	TACACCAGTT	AATTAATTCT	5520
TCTAAACGTT	CTTTTGTG	TGTTAAATGT	AAATAACCCA	TCAAGGATTC	GAAACCGGTC	5580
GCTACCCGAT	AGGTCGTCAC	ATCAGCATTT	TTGGCAAAGG	TGTGACTTTT	AGCATTCCGA	5640
CCGCGTTTGT	ACATCACTTC	TTCTTCTCC	GTTAGCATTT	CTCCCGCTAA	CATCGCTTGC	5700
ATTAAGCTTG	CTTGGGCTTT	TGCTGACACG	TAATGTGTCG	CCATGCGATG	CAAAACATTC	5760
GGTTTCGTTT	GTCCCAAGGA	AACTAAATAA	TCACGAATGT	AAATTTTATA	AATCGCATCG	5820
CCAACATAGG	CCAAAGCTAA	ACCGTTTAA	TGTTTGTAA	CTCTCATTCA	CTTCTTCTCC	5880
ATCTCGTTCC	TTGCGGAGTA	TCTTCTAAAA	CAATGCCTTG	TTCTTTTAA	AAATCGCGGA	5940
TTTCATCACT	TCGGGCAAAA	TCACGGTCTT	TCCGTGCTTG	GTTCCGTTCT	TCAATTAACG	6000
CATCGACTTG	TGCATCCACT	AATTCTTCGT	TTTTAAAGAA	TAAACCAAAA	ATAGCTAACC	6060
AGCCACTGAA	AAGCTTATCC	ATCGCCACTA	GAAGTGGTTC	CGAAACTGTC	GCTTGTTTAC	6120
TGTACTGATT	CATCATCTTC	GCTAATTCAT	AAACGACCGT	AATCCCGTTA	GCTGCATTGA	6180
AATCATCGTC	CATTTCCGTA	ATAAAACGTT	GTTCTAATTC	GTTTAATTGA	GCCAAATAGT	6240
GTTTCATCCTC	TGGCAAACCT	GCGACCGCAT	CTGCTTGTCT	AAAACGTAAA	TTTTCAAACG	6300
CATTTTTCAA	TTTTTGATAA	TTGGCTGCTG	CCTCTTTTAG	AGTGCTTTCG	CTATAACGAA	6360
TCGGGCGACG	GTAAGGGTG	GTGGACATAA	AGAAACGTAA	AATTTGCGGA	TCCACTTTTT	6420
GAATCATTTTC	ATGAACAGTA	ATGAAGTTAC	CTAGGGATTT	ACTCATTTTT	TCATCGTCTT	6480
CACCAATTGT	GACATAGCCG	TTATGCATCC	AGTAATTGGC	AAAGGTATGG	CCTGTTTTGG	6540
CTTCGCTCTG	TGCAATTTCA	TTCTCATGAT	GTGGAAATTC	TAAATCTTGT	CCACCACCGT	6600
GAATATCAAT	CGTCTCTTCC	AAGTGTTTTG	TCGCCATGAC	AGAGCATTCA	ATATGCCAGC	6660
CGGGACGTCC	TTTCCCCCAA	GGAGAATCCC	AAGAAATTTT	ATCTTCTTTG	GCACTTTTCC	6720
AAAGAGCAAA	ATCCAATGGG	TCTTCTTTTA	ATTGTTGTTT	AACGCCAGTT	CGTTGACTTG	6780
CCCCCACTTC	TAATTCATCA	ATCGATTGAT	GGCTTAATTT	ACCATAATTA	GGGAATTTAC	6840
GGGTGCGATA	ATACACATCA	CCTGCGACCT	CATAAGCAAA	ACCTTTTTC	ATTAAAACCTT	6900
CGATGAAAGC	TAAGATATCA	GGCATGTGGT	CCATCACCCG	AGGATGTAAC	GTTGCTGGTT	6960
GGACATTC	TGCTTGCGTG	TCTTCTTCAA	AGGCTTTGAT	AAAACGCTCC	GCCACTTCCG	7020
GCGCAGTAAT	TTTCAATTCT	TTAGCTGCCT	TGATAATTTT	ATCATCGACA	TCTGTAAAGT	7080
TCGAGACATA	ATTCACCTCA	TACCCACGAT	ATTCAAATAA	ACGACGAATC	GTATCAAAGG	7140
CGATCGCACT	GCGCGCATT	CCGATATGGA	TATAGTTATA	CACGGTTGGT	CCGCAGACAT	7200
ACATCCGAAC	TTTACGCGCC	TCAATTGGCG	TAAATACTTC	TTTTTCTCTG	GTCAATGTAT	7260

TATAAATTTT AATCATGCCC TTTTCCACC TTTCTCCAT TTAGACGAAC GACTTTCGCA	7320
GGAACACCCA CAGCTGTGGC ATCTGCTGGG ATGTCTGAAA TAACAACAGC ACCCGCACCA	7380
ATTTTGAAC GTTCACCAAT CACCACTGGC CCTAAAATTT GAGCTCGTGC CGAAATCATC	7440
GCTCCTTTTT TCACTGTTGG ATGGCGTTTA CCCGTATCAC GACCGGTCCC ACCTAACGTC	7500
ACTCCGTGAA ACAAACAAC ATCTTCTTCT ATTTCCGCGG TTTCCCAAT CACAATGCCC	7560
ATGCCGTGAT CGATAAAAAC TCCCGCTCCA ATTGTGGCTC CGGGATGAAT TTCTACGCCC	7620
GTTATGAAAC GCCAAAATTG TGCATGAATC TTCGCCAACA AAAACAACCTG ATGCCTATAT	7680
AAGAAATGAG ACAAGCGATG CCAAAACAGG GCGTGCAAGC CAGGATACGT TAAAAGAACT	7740
TCTAGTGTGG TCCGCGCTGC AGGATCGTTT CTTTTAACAG CTGCGACGCT ACGTTTCAAC	7800
CAACTCATAT GGTGAAGTCC TTTCTATTAT AAAAAATAA AAAACGTCTT TTAAGCAACA	7860
AAAAAACGTT GCTTAAAAGA CnCCCAtGCG TGGTTCCTACT TTTCTTACA AGAACTAGCT	7920
CTTGCCTCGA TACAGTTAAC GCCTGCCGAC GTCTTCGGTT ACTTTTTTCA CCAAAGCCAC	7980
TCATAGATGC ATTTGACAA AAGATGCTCG CTGTTCTCAC CACCCACAGC GTCTCTGAAA	8040
TCCATCTGAT TGTCTACTTC TTCTATTCTT TGTGTTTTTA TAACACTTTA TTTAAATGAT	8100
CTAATGCTTT TTCTTTGCCT AATAATTCGA TGGTTTCGCC TAATTCAGGG CCATGCATTT	8160
GTCCAGAAAC AGCAATTCGA ATTGGCATAA ACAAGTTTTT GCCTTTAACG CCGGTTTCTT	8220
TTTGGACGGC TTTAATCGCT GCTTTAATAC TTGGGACATC GAAAACGTCC AAGGCTTCTA	8280
ATTGTTCTTT AAACGCATGT AACACAGTTG GCACTGTTTC TCCTTGCAAG AATTCTTTCTG	8340
CAGCATCATC TAAAACTGGA TGTTCAATTA AGAACAAGTT TGATAGTTCC ACAATTCAG	8400
CTGCATAACT CATTTGTGGT TGATACAAGC TAACTAATTG TTTTAACCAT TCGATTTTTT	8460
CAGGAGATGG ATTTTCTTCA ACACGACCGT CAGCGACTAA ATAAGGAATA CACATATCTG	8520
TTAGTTCACT TAAGTCCATT TGTTTCATGT ATTGATTGTT AATCCATTCT AATTTTTTCG	8580
CATCAAAGGC CGCTGGTGAC TTGCTTAAGC GTTGTTCATC AAACAATTTA ATCAAGTCTT	8640
CTTGTGAGAA GATTTCTTCT TCGCCAACAG GAGACCATCC TAGTAAAGCA ATAAAGTTGA	8700
ACATGGCTTC TGGTAAGTAA CCTAATTCAC GGTATTGTTT AATAAATTGT AAAATTGTTT	8760
CATCACGTTT GCTCAATTTT TTGCCAGTTT CTGAATTAAT GATTAAGGTC ATATGGCCAA	8820
AGACAGGTGG TGTCCATTCA AACGCTTCGT AAATCATTAA TTGTTTTGGT GTGTTGCGAA	8880
TGTGATCATC GCCACGTAAA ACATGAGAAA TTTTCATTAA GTGATCATCT ACTGCAACGG	8940
CAAAGTTATA CGTTGGCATC CCATCGCGTT TTTGGATGAC AAAATCGCCA CCAATGTTAT	9000
CTGACTCAAA AGTAATTTCA CCTTTCACGA TGTCATCAAA TTTGTATTCT GTATTACGTG	9060
GCACACGGAA ACGAATCACT GGTCTAGAC CAGCCGCTTC TTTGGCTGCT TGTTCAGAAG	9120
GTGTTAAATT CGCACAAGTG CCTGCGTAGT GAGGCATTTT CCCACGAGCT TGTTGTGCTT	9180
CACGTTCTGC TTCTAATTCT TCTGGCGTAC AATAACACTT ATACGCACGA TTGCTACTTA	9240



ATAATTGATC AATCAATGGT TGGTAAATTT CTTTACGTTT TGATTGACGA TATGGACCAT	9300
ATTCTCCCGG ATGAGCAGGT GATTCATCCC ATTCCATCCC TAACCATGCT AAGTTTTCCA	9360
ACTGACTTTT CTCGCCGTCT TCGATGTTCC GTTTTGGTTC GGTATCCTCG ATACGAATGA	9420
TAAATTCTCC GTCATTGTGG CGGGCAAATA AGTAGTTAAA TAATGCTGTT CTTGCATTCC	9480
CAATGTGCAA GTGTCCAGTT GGAAGTGGTG CGTAACGCAC ACGAACTTTC GTCATGTTCA	9540
TTCTCTTTT CATTTTTTAA AAACAGACGG TTTAACGTCG TTTTCAGTTA TTTTCCAAT	9600
AAAGCAACGG CAATTGCGCC GATTCCTTCT TTTTACCAA CAAAGCCCAT CGTTTCCATG	9660
GTCGTTGCCT TAATGTTTAC TTGTGTGACC GCTAAGTGGC AACTCGCGGC TAGATTCTTT	9720
TTCATCTCTG CCAAGTATGG GCTCATCTTT GGCTCTTCCG CTAAAATTGT ACAATCAATA	9780
TTGCCAATAG TAAAGCCCGA CCGCCCAACT TTTTCATTTA CTTTTTTTAA TAAGTTTACA	9840
GAATTCGCAT TTTTGAAGT CGGATCGGTT TCGGGAAACA GTTGCCCAAT ATCACCAGA	9900
CCAGCGGCAC CAAGGATTGC ATCAATAATT GCATGTGTTA AAACATCGGC ATCTGAGTGA	9960
CCTAACAGAC CTTTTTCATA AGGAAGAGTA ACCCCACCAA TAATTAAAG TCGTTGCGGA	10020
ACTAATTGAT GAACATCAAA ACCTTGACCG ATTCGAATCA TACTAAGTCC TCACTTTGCT	10080
TAATTTTTAT CGTTTTTAGG ATGGTTATCC TTGCGATCAT TAATCCCGCG ACCCGCGTGC	10140
GCCGGTTTGG CAAAAATCAT ACGCCAGCT GCGGTTTGT AAGCACTTGT GACGACCACT	10200
TCGATTGCT CATTCATGTA ATGCTGGCCA TCTTCTACCA CAACCATTGT GCCATCATCT	10260
AAGTAAGCGA CCCCTGTGTG ACGCTCTGTT CCGGCTTTTA CAACCATCAC ATTCATCGTT	10320
TCACCTGGAA TCACCACTGG TTTCACAGCA TTAGCTAAAG CATTAAATGTT TAAAACAGGC	10380
ACATTTTGAA ATTCAGAGAC TTTATTTAAG TTGTAGTCAT TGGTCACAAC CACGCCATCT	10440
AATAATTTG CTAATTTGAT TAATTTACTG TCTACTTCAG AGATATCCTC GAAATCTCCG	10500
TCATACATCT CAACGGAGAT ACCTTCTTCT TTTTGTAAG CATTTCACG CTGGTTCCT	10560
GCCCTTTAAA ACTAATAATT TnAACTGTAC CGCCATGCC GACAACACGT CTCGGAATAT	10620
TAATTAAGCC ATAAC	10636

## (2) INFORMATION FOR SEQ ID NO: 75:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6531 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

GAAaGGGTCT TTTTTCGAGT AGACGTTTGA TnTTAGCGTA TCCAAGAAAG GAAAACCCTA	60
ACATAAAaAT TTTTATTTA CGAaGGAGAC CGATTATAT GGAAAAGAAA GAATTCACA	120
TTGTAGCAGA AACAGGAATT CACGCACGTC CAGCTACTTT ATTAGTACAA ACTGCAAGCA	180

AATTTAACTC AGATATTAAC TTAGAATACA AAGGTAAATC TGTTAACTTA AAATCAATCA	240
TGGGCGTTAT GTCTTTAGGC GTTGGTCAAG GTTCTGACGT AACAACTACT GTTGATGGTG	300
CTGACGAAGC TGAAGGAATG GCAGCAATCG TTGAAACATT AAAAAAGAA GGATTGGCTG	360
AATAATGTCT GAGATGCTAA AAGGGATTGC CGCAAGTGAT GCGGTTGCTG TTGCTAAAGC	420
TTACCTGCTA GTTCAACCTG ACTTGTCTTT CAACAAAAC TCCGTAGAAG ATACTGATGC	480
AGAAGCAACT CGCTTAGACG ATGCTTTAGC AAAATCTACT GAAGAGTTGC AAGCAATTCG	540
CGACAAAGCA GCTCAAAGCC TTGGTGAAGC AGAAGCGCAA GTTTTCGATG CTCATTTAAT	600
GGTTTTATCT GACCCAGAAA TGGTTGGTCA AATCAAACAA AATATCCAAG ACAACAAAGT	660
CAATGCAGAA GCTGCCTTGA AAGAAGTTAC GGATATGTAT ATTGGTATGT TTGAAGCAAT	720
GGaCGACAAT GCTTACATGC aAGAGCGTGC aGCAGATATT CGTGACGTTG CCAAACGTAT	780
CTTAGCACAT TTATTAGGTG TGA CTCTTCC AAATCCTTCA ATGATTAACG AAGAAGTAAT	840
CGTGGTTGCC CATGACTTAA CACCTAGTGA TACTGCACAA TTAGACCGTA CGTATGTTAA	900
AGCCTTTGTG ACAGACATCG GCGGACGTAC TTCTCACTCA GCCATCATGG CTCGTTCTGT	960
AGAAATCCCT GCAATCGTTG GTACAAAAGA AATTACTGAC AAAGTAAAAG CAGGCGATAT	1020
TTTAGCAGTG AACGGAATCA TTGGGGATGT TATTATTGAC CCAACAGATG CTGAAAAATC	1080
TGAATTTGAA GCAGAAGCAA AAGCTTATGC AGATCAAAA GCAGAATGGG ATAAACTAAA	1140
AAATGCAGAA ACAGTGACAG CTGACGGCAA ACATGTTGAG TTAGCTGCAA ACATTGGTAC	1200
ACCAAAAAGAT TTAGAAGGCG TACACAAAAA CGGCGGCGAA CTGTTGGTTT ATATCGTACA	1260
GAATTCTTAT ACATGGATTC ATCTGACTTT CCAACTGAAG AGGACCAATA CCAAGCGTAT	1320
AAAGCAGTTC TTGAAGGAAT GGAAGGAAAA CCAGTTGTGG TTCGTACAAT GGATATTGGT	1380
GGGGACAAAG AGCTTCCTTA CCTAACATTA CCACACGAAA TGAACCCATT CTTAGGTTAC	1440
CGTGCAATTAC GTATTAGCTT ATCAGAACTA GGTGACGGCA TGTTCCGTAC ACAAATGCGT	1500
GCATTATTAC GTGCGTCTGT TCATGGTAAC TTACGTATCA TGTTCCCAAT GGTGCTACT	1560
TTGAAAGAAT TTAGAGCAGC GAAAGCAATC TTTGAAGACG AAAAACAAAA ATTAGTCAAC	1620
GAAGGTGTTG AAGTTTCAAA CGATATCCAA GTAGGTATTA TGATTGAAAT TCCAGCAGCT	1680
GCCGTATTAG CTGATAAATT TGCCAAAGAA GTTGACTTCT TTAGTGTTGG AACTAACGAC	1740
TTAATCCAAT ACACAATGGC GGCAGACCGT ATGAACGAAC GCGTTTCTTA CTTATACCAA	1800
CCATACAACC CATCAATCTT ACGTTTAATT AAAAATGTAA TTGATGCAGC ACACGCTGAA	1860
GGTAAATGGG CTGGTATGTG TGGTGAAATG GCGGGCGATC AAACAGCTGT TCCATTACTA	1920
TTGGGTATGG GCTTAGATGA GTTCTCAATG AGTGCTACAT CTATCCTTAA AACACGTAGC	1980
TTGATGAAAC GTCTAGACAC AACTAAAATG GCTGAACTTG CTGACCGCGC ATTAAAAGAA	2040
TGCGACACGA TGGAAGAAGT CTTTGCATTA GTTGAAGAAT ATACAAAATA ATTAAGTGAT	2100
TATGAAAGCA CATCGACAAT TGTCGATGTG CTTTTTTATT GTTGCATTTC ACGTATTTTT	2160

TTCTTGAAAT AATAAATCGT GCTTCCCAA ATGAATAGAA AAACCAGCGT GATGCTACTC	2220
ACCAGTTTCC AAGGGAAATT GGTGTGTTTT TCAGGTGGGA GGTGACGGC GGTAGCATTG	2280
TATTTTTTAG CTTCTTTTTC TTGGATGACT AAGTGTTTTT CCCACGAATA GTCTTGATAG	2340
CCATCGTTTG CTTTGATTTT TACAGTATAA TTTCCAGGAA CAAAAGGTTG TTCCTTTAAG	2400
TCAATCCGAT AATTAAAATT TGTATTAGGA GCCATTGTTA ACGACTCTTG GTGATTTTCG	2460
TAGCGAGGTT TGGTTTCTTT TTCATAATAC AAAGAGGCAT CGACTTGTA TTTTAAATC	2520
ATAGCCGCTT GTTTGTTTTG GAGATTCATT AATACGCGT TACGACCGTT AGCTTGATCG	2580
GTTTTTACTT CATTTAAAGA TAATTCTGGC TGAACAGGTT CATCTGTTTC ACTAAGTAAG	2640
ACACCGACTA CATAAGAAAA ACGATTATTG ATGGCCACAC CGTTTCTGA GTGAGCAGGT	2700
TGTTCATCTG ACTTTTGTTT AAAGTAGAGT CCTCCTAAAA CGTACCCATT AAATGGTTTT	2760
TCCGGTAAAG TAAGCTGACA TTCAACAGTC TTGGTTGAGT GTTTAGGTAG GTGGATTCT	2820
TTTGGCAATT GCGCAATCTC ATTAAAAGTA AAAGGTGCAG ACGGATCTTT TTTTGTTTCA	2880
GCATGAGAAT AATCTGCCAA GCCATTATCA TTAGTAATCG CTGCATTGGC TGAAGCTAGC	2940
ACGGTAATGT CGTTGGCTGT TTGATTCGTT AGTTCGATT GTAGCTTTTG TGTTTGTTT	3000
GGAGCGACAC GTAAATCAAA ATAACCTGTC TCTTTGGTTC GTTGATTGTC GGGTAAATG	3060
GCTTTTACAG AAAAAGTCAT TTCGCTTGCT TCTACAGAAT AAGAGCCAAT AAAAATAAA	3120
AAAATGATGA CGCTAATAGC ATGGAGCCAT TTAAAAAGCT GCTGATTCAT AGTGTGCACT	3180
CCTTTTTGAC AATTTTTTAT AACAAAGTAT TGCATGAAGT TGTTTCATGC AATACCTTGT	3240
TTGTTTTCTT TTAAAGTGGT GTATCATCCA AAATCCAAGT AAGCGTTGTT TTATATTGTT	3300
TAGCTGCAAC TTTTTCGTT GTCGCAGGAA CAGTTAATTG AATGCCTTGA GCAGCGGTcg	3360
sTCCGCTACC AAAAGCTAAT GTCCAAGTGC CCATTCCTGT TGAAAGAGCG GCAGTCGCCA	3420
CAGGAGAAGT GGCGGCACCA GTTGATCCA AGGTACTGG GACCGTTAAT GGCGCCAAAG	3480
AAAGTAACTG TGTTGAATTG GGTGTTGCTG CTAAAAGGT CAAGGATGCA TTATCCAAAA	3540
CAGCGGGTGT TGAATCACTT GTAGCAAATT GCGCACTCTG TTAACTGAT AATTTCCAGC	3600
CAAGATTTAG ACCACGTTTG TCAGTTACTT GAACATAGTT TGGCACGCTA ATTAAATCGC	3660
CAGTACTATT TTGCACTTGA TCCAGTTGTG CCGAATAGAT CGCTGTTCCG GCTTGAATTT	3720
GTTTTGATCC AAAATGGATA TTTGAAACAT AGTCAATACT AAGGGGTCCG GCTGTACCAG	3780
GTTGATGAGG ATCAGCAGGA TTAGGTGTCA CAGGCTGAGA AGGGTTCGTC GGGTTGACAG	3840
GTGGCGTGAC AGTATTATCA AGAGCAAACG TAATATCAGC GTTGATTGA ATAGAAGCGA	3900
CTTGAGCTGC TTCAACTTTT GAACTATTTA AAAGCATAAG AGGCGTAgcA ACGAACAAAA	3960
TGCTTGAAAT GCGCACGATC TTTTTCATTT TTTATTCCTT CTTTCTTGTT TATTTTTAAC	4020
TTTTAAAAAA GCTAAATCCT AATTAGAGGG GTGTATCAGA AAGAATCCAT TGTAATTCGG	4080
CAGTGTAAGA TTCTTTGACC GTTAAATTAA GGTCAGGAGC AACGCTTAAA AAAAGTCCAT	4140

TTTGTGGTT CCAAGGAATT TCTACATTGT TATCGTCCGT TGGGTGTTT TGATAACTT	4200
CGGTTGCTAA AGAGGAGAGA GGAAAATCAC TGCCATTTTT TCGATAAATG AAGGCTTGCT	4260
GGATCGTTTT ATTTGTAGAA GTTGAATGCA TGGGCTGACT CTCTCTTACC AACAAATGCC	4320
ATGGTGT TTTT TGCAACTCGA CCATCGGCAA TGAGGAAAGG AGCGCTGATG GTCGATGGTG	4380
AGTATATGGT TCGTTTGCTG GGGAAAGGTA AATTTTTTAAA GGAAACTGTA GTTGGCATCT	4440
GCACAAAACG GAGGGTACCG TTTTGTATAG GTAAAGATTG TTCTTTTAGTG ATTGGCSTAA	4500
TTGcGTTCCG CCACCATCAT TGCCAGTAAC CGTGCTAGAA AAATAAAGAA ATTGGTTTAT	4560
CCATTCTTCT GTAACCGTTG CTGTATATTG TAAGGTGAAT TGGTATTGG CTTCCTTAA	4620
ATTATCTGCT GGCAGCGAAT TATGATAACT CCAATACGGC AACCAGTGG AACTATCAGT	4680
TAAAAGTTGA AGCCCAGCTA TTGGCGTACT TAAACCTGCT GCTGATTGTT TTGTTTATGT	4740
AGCAGCTGAA AGTTGGGACA GTCCTTTAGG CACCGTTGCT CGTAATTGCT GATTCAATCCa	4800
CGTTGCTGGT TGATTGGTTA AACGTGTGGT TAACGTAAAA CGAATGATTT GACCACTTAC	4860
GACAGAGTTT GCAGCTTCCC AGAGCCCTTG TGCATTTTTT CGCTCAATGA TTTGCGTAAC	4920
CGTTAGTTGA GGCACAGGAT AAGCAACTGT TTGAGAAACA GTAGAACTGG TTAACAAAGG	4980
CAGGCCAGTG CGATTGCTGA GACCTTGGAC CCACAAGCTG ACCGTTTCTC CCCCTTGGAA	5040
AGTGGGTAAA TTAGCAAATT GAAAATGGCC ATTGGAATCT AAGGACGTGG TGCAATAGA	5100
ACCACCATT ATTTGTACTT TAACTTGATA GGTTCGGTTA GATACTTGGG AATCTTTACC	5160
TGAAAGTTGA GTTGAAGCCG CATTGGCGAC TACGTTTGTG ACTGGTTGGA TTGACCAATT	5220
GACACTTGG TCAGCTGAGG GGCTGCTTGG AATATAGTCG GTTGACGTTT TTGAACGAGC	5280
AAAGACAATT GTAGAGAGCG TGTCGCCATA TTCTAATGGT GAAGCTAAGG ATGTCCAGAA	5340
ATTTCCGGCT GCATCGACAG AAATCCCTTT CACCACAATG GGGGTAGTGG TGCCATCATG	5400
GCGGTTCAAG GTTAATTGAA CAGAGTAGTT ATAGTTTGGG TCTTGTAATT GTTGAATGGT	5460
CCCTGTTACA GTAGTGCCA GATGATTTGG TTGAGTTAGT TTGGCTGGTG TGACTTGAGG	5520
AACGGGATAA TTACCTAAAC TAAGTTGAGC TTTGCTAATT ATGCCTGTTT TTGTATTATT	5580
CCCAATGATA ATATTAATTG CGCCTAATCC aGTGCCGACA TCTGTAAAGG TTCCTCGAAT	5640
CGGATAGTCT TTGCCGTTAA GAGCAGGAGG GATTGAAATT CCTCGTTTAT AAAGTGCGCC	5700
CACATCAAAC GAAATGGCGG CAGACCATTG GGTTCGGGA AGTAACACCA AGGTGCTACT	5760
TGGAAGATC AGATAAACAG AATGATAAGT GCTACTATAA GAGAGCTGAC GCCGTTAGTT	5820
GAAGTGTGAA TATCTTGAGT GGCTGTTCTT GGATAACTGA CAGTCCCTGT TAAAGAGTC	5880
AGAAAATCTT GTTGCTTTAG AGTGTCGTTA TCTAACTGGC TTGAAATTTT TGTAGGAAGT	5940
TGAATAGCAA TAATCGGTCG ATCATTAAAT GCTAAGCCGA TAAGTGTCAC ACTGTATTCA	6000
TACGAAAGTG TGATAATATC TTGGTTATAC GTTCCCGTGC TTAAGGTACT TTTTAAGGTG	6060
CTATTACCAA CAGAAACACT AGCTGGCGTA GGTGCCATAG GTTCATCCTT TTCAGAGGAA	6120

GCATGGGTTT GCTCGGTTTC TTGTTCCACG ATATCGGTTG CTTGAGTTGT TTGTTGTTCT	6180
GTTGAAACTG AAGTAAATGT TTCCTAAAA GTAGAGGCGT GAGTTGTAGG GTTCAGTGTT	6240
TCAGAATGAG CTATCAAAGG TTGAGTCAGC GATAAAAAAA GCCATGATAT GATGCAAGAA	6300
AAGATTCCCC ATACGAACTT ATTCAAAACC ATCCTCTCCT TTTTGTATAT TATATGTATA	6360
ATGACGAAAA TATATAATGT GTAAATATTA AGTACATACA TTCTATAATG GGCCATnCAT	6420
AAAGTAAAGC GATTACATCC GATGGTGnAC AGGATATTTT GTTCACAATC TTTTAGTGTT	6480
TTAAAGAGGG AAATTTCTCG AAATTCTATG CGTTTTnAAT CCGGGAAAGA T	6531

## (2) INFORMATION FOR SEQ ID NO: 76:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6399 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

TTTTCGGGAA AGAAACACCT ACCATTTTGT ATGTTTGGG CAGGGAATAG GAGGTAATAC	60
ATGGAATGGA GTTTCCTTG GCTATCAATA GATGGCGACA GAATGTACGA TGATAGCGAC	120
TTCTCGAAAT TTTTGAGGG TCTTTTTTCT TATGGCGTTT CTCTAACAAC AGCTAACGCA	180
TTAAAGGTTA CTGCCAGTCC TAATGGCGGG ATGAAAGTAC AAGTAGATTC AGGCTATGCA	240
TTTGTCGGTA AAGTCTTCTT GAATAGTACT ACAAAGCTT TAAGTATTGA TGATAGCAAGC	300
AGTATGCAAG ACAGGACGGA TAGTATTGTG GTTCGCGTGG ATAAAAGTGT GCGAGATGTA	360
TTTTTAGCTG TTAAGAGAA CGATACGACA GTAACCCGTA CATCGGACGT TTACGAGCTA	420
CAGTTAGCGA CAATAAGAGT GCCCGAAAT GTTCTAGTA TTACAGGCGA TTAATCACA	480
GATAAACGGG CAGATACGAA AGTATGCGGG TATTCTTCGC CGTTTCAAAA GGTTAACGTT	540
TCAGGGTTAG AAGAGCAATA CGGTACAATA CTAAAAAAA TTGCGGAAGC CAACAAGACG	600
AGTTATGAAA AAATCCTAAA TGATTTTAAA AACTATGTTG CAAAAGCGCA AACCGATATG	660
GATTCTAATA TTGAAGAAAT TATCCAATCA GGTAATGGAA AGGTAAGTGC TTTGATGTT	720
TTAATTCACG AATGGTTTGC AGCTTTAAAA AATGAATTAG ATACAAATCA AGCATCAAAT	780
TTACAGAATC AAATCAATGA AATGAAAGCT ACTGAAGAGT TACCAGCTAT AGAACATAAT	840
TTACTCGGTT ATCCTAATGT ACAAGTTTGT TATTGGGAAT ACGGTATTGG CCTATCAGGA	900
TTAGCTAATG AGCCAACAGG TCTAGGCGGT AGCAATGTGA AAAAGATTCC TCACAGTGTA	960
GAATATCTTG ATTTATTCAG TTTCAAAGTT AAAGTGCCAA TGAACTTTAA ATTGGTAAAT	1020
CCAACAGTAA CAAAATAGA TAGTCGAACT ATTCGCTTTA TTGAAGCATT TAAAGTTATA	1080
GAAATTAAAT TTTAGGAGGA AAAGAATGTA TACATTTAAA AAAGGTGATG CAGACTACCA	1140
AGTCATGCTA AATGAAAATT TTAAAGAGAT AACTGATTCT TTCGAAGATG GTTTATTGTT	1200
CAGAAAAAAT TCTAAATAT TAGATTTAAA CGATGCTATA TTGCCAGGCA TTTATTCAAT	1260

CCCAGCTACA GGAGTTGATA ATAAACCTTT ACCTAATTCT GGAAGTTTAT TCGTTTGTAA	1320
AGACCCAGGA GGGGTTAGAC AATTATTTCA GACAGAAAGA ACAATATTTA TTCGACAATT	1380
AGGGGGAGTA CCTTCATCTT GGaCGGATTG GaAAAAAGTA GCTTTTGAAA AAGAACAACC	1440
TTTTGAAGCT TGGTATTCAC CAGGAActAA CCATGCTGGA TTCAAAAATA AGGCAAGATA	1500
TAATTTAGGG CCAGAATTTA GCAACATAGG CCAGCGACTT GGGTTGCCTA TGAAAAGTGA	1560
ACCGTTGGAG TGGAATAGTG GGCATGGCA AGCAAAAGTT CTTAGAGACT GCAAGCTAAA	1620
TATAAGCGGA ACTGTAAAAT ACCATGTTGG TAGTTCGAGA GGTGTTCTCT ATGCTTATAC	1680
TCATATAGAC AAAGGCCTTG ATGAAGGCGT AGGTGACTTA GGTATTGGAT CAGCAGTCGG	1740
AGCTGTTGGC GGCTTGGATT ATAAAAATGT CGCAGCTTTT GATTTAAACG TTACACTAAA	1800
AAAAGTGAG TATCTTGCCT TTCGCTTAGA ATTAGCAGCA GATAAGCAAC TTGATTATAC	1860
TCAATTATCT TCTATGCATA TCACAGAATT AGTATAGGAA CTGAATTTTA AAATAAAAAC	1920
CGTTTAGCAA AAAAGCTAAG CGGTTTTTAT TTATAGAGAG AAGGAACTTT ATGTGAAGGA	1980
TGAAGTATT CAAGATGTTG TAGAGCGTTT AGTGCGTATT GAGACAAAGC TCGATAACTA	2040
CGAAGTGTTA AGAGAAAAGA CAGAAGATGC TAAGGAAAAG GCTGATTATG CTTTCTCTAT	2100
TGCTAAAAAT AATGAAGAAG ATATCAAAGA AATAAAGGAA AATCAGAAAT GGTCTTGGCG	2160
GACAATTGCG GGCATAGGGG TGTCAGTAGT TGTCTATTTA TTTACAAAAT ATTTAGGAGG	2220
TGTATAAGAA TGATTCTACC AGATAAGTAC TACAAAATTA TCAAATGGGG CGTGCTAACA	2280
GTGCTACCTG CAAGTTCTGT TTTGGTTGCC ACGTAGGTA AAGCTTATGG ATGGCAGCAA	2340
ACAGATATGA CTGTTTTAAC TATTAATGCC ATAGCAACTT TTTTAGGAGT GGTAACAGGA	2400
GTATCAGCAT ATAATTTAAA AGACAAGGAG AAATAAAAAT GAAAAGAAA ATTTTAGCAG	2460
GAGCGCTTGT CGCTCTGTTT TTTATGCCTA CAGCTATGTT TGCCGCAAAA GGAGACCAAG	2520
GTGTGGATTG GGCATTTTAT CAAGGTGAAC AAGGTCGCTT TGGCTATGCA CATGATAAAT	2580
TCGCTATTGC CCAGATTGGA GGCTACAATG CTAGCGGTAT TTATGAACAA TACACATATA	2640
AAACGCAAGT GGCAAGTGCT ATTGCCCAAG GTAAACGTGC GCATACCTAT ATTTGGTATG	2700
ACACTTGGGG AAACATGGAC ATTGCGAAAA CAACAATGGA TTACTTTTGT CCACGTATTC	2760
AAACGCCTAA AAATTCCATC GTTGCAATTAG ATTTTGAACA TGGAGCGTTG GCTAGTGTTT	2820
CAGATGGATA TGGAGGATAT GTAAGTTCAG ATGCCGAAAA AGCAGCAAAT ACAGAGACAA	2880
TTTTGTACGG TATGCGCAGA ATCAAACAGG CTGGCTATAC TCCAATGTAT TACAGCTATA	2940
AGCCATTTAC ACTAAATCAT GTAACTATC AACAAATCAT CAAAGAGTTT CCTAACTCTT	3000
TATGGATTGC TGCGTATCCT ATCGATGGTG TGTCACCATA TCCATTGTAT GCTTATTTCC	3060
CAAGCATGGA TGGTATTGGT ATTTGGCAAT TCACATCCGC TTATATTGCA GGTGGTTTAG	3120
ATGGTAACGT AGATTTAACA GGAATTACGG ATAGTGGTTA TACAGATACC AATAAACCAG	3180
AAACGGATAC GCCAGCAACA GATGCAGGCG AAGAAATTGA AAAAATACCT AATTCTGATG	3240

TTAAAGTTGG CGATACCGTC AAAGTGAAAT .TTAATGTAGA TGCTTGGGCA ACTGGGGAAG	3300
CTATTCCGCA ATGGGTAAAA GGAAACAGCT ACAAAGTGCA AGAAGTAACT GGAAGCAGAG	3360
TATTGCTTGA AGGTATCTTG TCATGGATTA GCAAAGGTGA TATTGAATTA TTGCCAGACG	3420
CAACAGTCGT CCCTGATAAG CAACCAGAAG CGACTCATGT GGTACAATAC GGAGAAACAT	3480
TATCAAGTAT TGCTTATCAA TATGGAACAG ACTATCAAAC GTTGGCGGCA TTAAATGGAT	3540
TGGCTAATCC AAATCTTATT TATCCTGGTC AAGTTTTGAA AGTCAATGGA TCGGCAACAA	3600
GTAATGTCTA CACGGTTAAA TACGGCGATA ATTTATCTAG TATTGCAGCA AAACCTGGCA	3660
CTACTTATCA AGCTTTAGCT GCATTAAACG GATTAGCAAA TCCTAACTTG ATTTATCCAG	3720
GTCAAACATT GAATTATTAA TAGCTTTAAT ATAAAATAAG GATACACTTA TTTAAATTTT	3780
TCTCGAGTCG CCGTCGCCAA GCGGCTCTT TTCAGGACCA TTAGCTCAGT TGGTTAGAGC	3840
AAACGGCTCA TAACCGTTCTG GTCACAGGTT CGAGCCCTGT ATGGTCCATA GCTAAAAAGG	3900
GCACGAAAAG GGCAAAAGTA GTAATACTTT GTCAGATTAT GTTGTTTTTT ATAAATTTTA	3960
TTAGTATTAT ATAGCATTAT AAATGCCTGT TTTACGGGCT TTTACAATTT TTGTGAAAAT	4020
TGGACCAAAT CCTTGGGATC CAATGGGCGC ATTAGCATAA AACAGAAAAA GAAGGCCCCC	4080
GCCGTCAGGG CGGGGGCCTT CTTTTTACGT TGGGCGGCGA ACTTCACGGA TTTCGCCCAA	4140
TGTTGTATAA TCATTGCCTG CTAAACGAGC AATCGGTTGG AGTTTTTCAG GTAAATGTA	4200
CTGATGTTCT TCATCGAAAA CGGCTGAATC AAAATAGAAA TCGGTAATTT CTACAATGAA	4260
TAAATCCGTA ATAATTTGAC CTTTCATGGT AGCAATAGGT ACATATTGAT GTAAGCGAGC	4320
TTCCATCCGA ATCGGCGCGG CTTTGATTGC TGGAACAGCA ACAGTCTGGC TGGCAATGGT	4380
TTCAATACCA AATGTGTCAA TTTCACTGAT CTCTGCTGCT AAAGACGCCG AAGTCTGATT	4440
CATTTGAGTA AGAACGTCTT CGTTAACGAA ATGGATGACA AGTTCTTTAG TAGCCAAGAT	4500
GTTTCGTGCC GTATCTTTAG GCTCTTCGTC TGGTCGTAAA ATGGACAACG TTGCTAAAGG	4560
AATTTAGCT GCCCGGCAT TAAAGAACT AAAGGGAGCA GCGTTGACGA CACCTGTAGC	4620
TGTATCTTGT GTTGTAACCC AAGCGATTGG CCGAGGAATA ATACTTCCTG ATAAAAATTT	4680
ATACGCTGTT TTTGCAGATA ACTCGTTACT ATTGTAATGA ATCATAGGAG CCTCCTAGAA	4740
AAAATACTAA TTTCTTTGAA CATTCGCATC TGAAGTATCG AAAGGACGCA CAAATTCCTC	4800
AATCCCTGCC CGTTTTGGTT CTAAAAATGG CGGTAATGAT AATTTTTTCG CAGCGTGTTC	4860
ATAGGTTTCA TCCTCTAAAA ATCCAGGACC ATCGGTTGCC AGTTCAAATA AAACATGTGT	4920
TGCTGCCAAG AAaTATTCAG ATTTAAAGTA AAAaCGTtCC ACAAAGCCAG AATTAGGAAA	4980
TTCCAAACGA TTAATTTTTT CAATCCAGAA ACGTAGTGCT TCCTCATCGG CAACGCGCAA	5040
GGCTAAGTGA TGGACATTGC CGTAGCCTTC TTGAGCCGCT GGCAAATCTT CCCGATGCTC	5100
TACGATGATG CTGGCACCAT TCCCACCTTC GCCAACTTCA AATAAATGAA AACTACCTTC	5160
AGCATCAATC AATTTGAACC CTAAACTTC AGTTAAACT AGTTGCAGAT GTTCGTAATT	5220

TTCGACAGTT ACAAAGGCTG GTCCTAAGCC GATTAAAGCA TGTTCACTAG GGACGTTAGA	5280
TTTTTCCAA GGCGTGCCGG CTGCAACACA TTGGTTCTTT TCATCAGAAA TAAGTTGGTA	5340
ACGTTGATTG TCAAAGTCTT CAAATTCTAA ATACTTTTTA CAAAACGTT CTTGAATTTT	5400
TCCATGATCA ATTGCATATT CATTAAAGCG TGATACCCAG TAATCCAAAG TAGCATCATC	5460
TTTTACACGA AAGGAAGTAC GAGAAATTGT ATTGGTTCCCT TTGGTTCCtT TrGGTAAGCC	5520
TGGGAAGTCA AAGAAAGTCA TATCTGTTCC GGGAGAGCCG GCATCaTCAG CAAAAAwTAA	5580
aTGATAkGTT TCAATATCGT CTTGGTTGAC AGTCTTCTTA ATTAAACGTA GACCTAAAAT	5640
ATCGGTGAAA AAGTGATAAA TTTTTCGGC ACTTGAAGTC ATGGCAGTGA CGTGGTGTA	5700
ACCTAAAATT TGTGTATTCA TAATAAAAAT CCTCCTAAAA GAAAAATAAT TTTTGTATTA	5760
ATTAACCTAC TAAAAGTAAG TTAAATGAT TAGTCCTAAA AATGCAAAAG AAATGTCTCG	5820
AATTCGAACT ATTATTTTTT AGTTGATTTT TTAAAAAAT AAGAATATAA TAGCCTGAGA	5880
TATTTTAATA CATAAACTTT CGTAAACATC GCCGGATGCT TGCGACTAAA GTGTTAATGA	5940
ACAGTTCTTA GGCCATAGCT GAACTGTTTG TTGGCACTTT TTTTGTGGTGG AAGGGAAGGG	6000
AAACAAGATG AGTAAAGCTT GGTTAAAAGT TATTTTGGC GCTTTTGTG AAGTGATTG	6060
GGTCATTGGC ATGAAGCATA GCACAACATG GTGGGAGATT TTAGGGACAG TAATCGCAAT	6120
CTTATCAGT TTCTATGCGC TGATTAAAGC AGGAGAAGAA TTGCCAGTCG GTACAGCCTA	6180
TGCTGTATTT GTTGCTTAG GTAGTGCTGG AACTATTGTG ACAGATCATT TTTTATTCCA	6240
CACACCATTG GGCATAGGAA AAGTTCCTTT TCTGTTGCTG TTATTAATAG GTaTAATTGG	6300
CTTAAAAATG GTAAGTGGrA ATAAAGAGGA GGAAACAAGA TGAGTTGGTT ATACTTAGTT	6360
ATTGCAGGCT GTTTTGAAAT TTTTGGCGTA GGTTCaATT	6399

## (2) INFORMATION FOR SEQ ID NO: 77:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3956 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

AACAGGAGGA AAAGTTGATG GGCATTCCAA TCACCATTGT TTCTGGCTTT TTAGGCGCAG	60
GAAAAACAAC CTTAATTAAC CATGCGCTTG CGCAATCACC GTTTCCTAAA GAAGAAATTA	120
TTATTATTGA AAATGAATTT GGTCAAACAG GCGTTGACCA CGCATTACTT TTACAAACCA	180
CGGAACGAAT TATTCAATTA AATAACGGTT GCATGTGCTG CAGCTTACGA GGAGATCTTC	240
TTGCTCTTTT ATCAGCAATC TTAGAAGTGT ATCAAGAAGA ACAACAACCC ATTGCCCAAG	300
TCATTTTAGA AACAAACGGGC ATTGCTGATC CGCAACCAAT CATTCAAACA ATTTTAACGA	360
CCCCGCACAT CAAAATCAT TTCTATATAG ACAGTCTCTT AACGGTCGTT GATGGGCATC	420



ATTGGCAACA ACAGCTTCAA GAAGCCGAAG CCATCAAACA ACTGGCCTTA GCAGATCGCC	480
TCTTTTTTTC TGTCAGGAA CCAGCAGACA GCACTCAGCT ACCAACGTTT CAAGAAACCC	540
TTCGTACCAT TAATCCTTTT GCTGATATTT TATCCTTTCA AGCCAACGAG CCTCTTTTGG	600
CTAGTGACTT TTTTCAATTA AATAAGTTTA CGGCAACACT TACAGAACAC GAAATGACTC	660
ATTCCCCGCA TGAGCACGCC CACGAGCATC ATCACCATAC CTTTCACTCC CTCACGCTAA	720
CGGCGTCCTC AGCCATCAAC GAACCTCTTT TCACTCGTTG GCTTGATTGG TTAATGTATA	780
CGCATCAGGA AAAGTTATAC CGTTTCAAAG GGATTCTCGC CTTACAGGAG CATGACTTGG	840
CCATTGCCAT GCAAGGCGTC AACCAGCAAG TGGCCTTTCA AATGACGAAT CAACCCCTC	900
AAGAAACGAC AATTGTTCTG ATTGGCAAAG AATTAGAGAC AAAAAAATT CAAGAAACCT	960
TTCAAACCCT CAACAAATA GCTACTCCGT AAATTAATTT TTTACGAAAT GGAGGTCCTA	1020
TATATAATGG AAAAACTGA CTTATCTAGC GCTTATCGTC GACTGAAAAG TCCCAACATC	1080
AAAACACGAA AACGCGCACT TAAATCATT AAAGAACATA AGCGCAATAA ACAAAGAAA	1140
ATTGCGTAAT TCACTAAAGG AGCACACCTA TGAAAAAATT TACTCTTCCC CTGTTAGCCG	1200
CCTTATCGCT AATCCTTTTC GCGCTTGCG GCAAAACAAA CACCTCTGAT AAAACCGCTG	1260
ACGGTAAAGA AAAACTATCC ATTGTCACGA CTTTTTATCC TATGTACGAT TTTACTAAAA	1320
ATATTGTAGG CGATGAAGGA GACGTCAAAT TGTTAATCCC TGCTGGTTCT GAACCACACG	1380
ATTATGAACC ATCCGCCAAA GATATGGCTA CCATCCATGA TGCGGATGTT TTCGTTTACC	1440
ACAATGAAAA TATGGAATCT TGGGTACCAA AAGCTGCTAA AGGTTGGAAA AAAGGAGCCC	1500
CGAACGTCAT TAAAGGTACC GAAAACATGG TCTTACTTCC CGGCAGTGAC GAAGACGGAC	1560
ACGACCATGA CCACGAACAT GGCGAAGAAG GCCACCACCA TGAATTAGAC CCGCATACTT	1620
GGGTTTCGCC TCATCGTGCC ATCCAAGAAG TCACAAACAT CAAAGAACAA TTAGTCAAAC	1680
TTTACCCTAA AAAAGCCAAA ACATTTGAAA CAAACGCAGA AAAATACTTA ACAAATTA	1740
CAGCCTTAGA CAAAGAGTTC CAAACAGCTT TGAAAGACGC TAAGCAAAAA AGTTTTGTTA	1800
CCCAACATGC TGCATTTGGT TATCTTGCCT TAGATTACGG CTTAAAACAA GTGCCAATAG	1860
CTGGTTTAAC ACCTGAACAA GAGCCAACCG CAGGCGCTTG GCAGAGTTGA AAAAATATGT	1920
CACAGACAAC CAAATTCGCT ATATTTATTT TGAAAAAAT GCCAACGATA AAATTGCTAA	1980
AACGTTAGCT GACGAAGCGA ATGTTCAATT GGAAGTCCTA AACCCGCTAG AAAGTTTGAC	2040
ACAAAAACAA ATGGACAATG GCGAAGATTA TCTTTCTGTA ATGAAAGAAA ACTTAACTGC	2100
TTTGAAAAAA ACAACAGATA CAGCCGGGAA AGAGGTTTCA CCAGAAACCT CTGAAAAAAC	2160
AGAAAAAACC GTGGCTAACG GATATTTCAA AGACAGTGAG GTGGCTGAGA GAACACTGAC	2220
AGATTACGCT GAAATTTGGC AATCCGTCTA TCCTTTATTA AAAGATGGCA CATTAGACCA	2280
AGTCTTCGAT TACAAAGCGA AACTGAAAAA AGATAAAACA CCAGCCGAAT ACAAACCTA	2340
CTATGATGCC GGCTATCAAA CCGATGTCGA CCACATCAAC ATCACTGATT CCACCATTGA	2400

ATTTCTGGTC AATGGCAAAC CACAAAAATT CACCTATAAA GCAGCCGGTT ATAAAATTTT 2460  
 AAACATATGCA AAAGGCAACC GTGGCGTCCG TTTCCTTTTT GAAACAGACG ATGCCAATGC 2520  
 TGGGCGGTTT AAATACGTCC AATTTAGCGA CCACAACATC GCACCAACGA AAGCCGCTCA 2580  
 TTTCCACATC TTCTTCGGCG GCGATAGCCA AGAAAGTCTG TTCAATGAAA TGGACAACCTG 2640  
 GCCAACGTAT TATCCAAGCG ACTTAAGCAA ACAAGAAATT GCCCAAGAAA TGATTGCGCA 2700  
 TTAAGCATTC AAAAGAGAGG AGGTCTGGGAC AGAAGTGTTT AACTCCGAGA AATAAGAAGA 2760  
 AATTTCCGAA AATTGTTCTT TAATTTTTGG AGAATTTCCG CTTATTTCCG AAGGAGTTGC 2820  
 TTCTGTTCCC GCCGTTTATC AGTTTTTGAG CGTGGAGCAA AAATCCAAAG TGATTTTTGT 2880  
 CCCACGCTCC TCTCTTTTTT TATTTAGTCA AGCACACCTG kTCTGTTtCA AACGTCGCTA 2940  
 ACamTGGCGC AcTTCATCCG TTGATTTkGT ATTCATTAAT TGTGCTCTTA AATCACTGGC 3000  
 ACCTGGAAAG CCTTTGACAT AAATTTTAAA GAAGCGATGC AGCCCAACGA TGGAGCGAGG 3060  
 CACCAATTCC GCATATTTGT CTTGCAAATC CAATTGTAAA CGTAGCAAGC CCAGCAATTC 3120  
 TTGTGGCGTA TGTGTTTTGG GTTCTTTTTT AAAGGCATAA GGATTTTTAA AAATCCCCCG 3180  
 ACCAATCATG ATGCCATCTA CACCATATTG TTCTGCTAAT TCTAGGCCCT TTTGACGATC 3240  
 GGGAAATATCC CCATTAATCG TAATTGTCGT TTGCGGTGCG ACACGGTCCC GAATAGCCAT 3300  
 AATTTGCGGA ATGACCTCCC AATGGGCATC CACTTTGCTC ATCTCTTTTC GTGTTGCGAA 3360  
 ATGAATGGAT AGATTCGCAA TGTCTTGCTC TAATAAATGC GTGATCCACG CCTCCATTTC 3420  
 CGCCATTTCA GTAAACCCAA TGCGTGTTTT GACACTAACA GGTAAGCCAC CTGCTTTGGC 3480  
 AGCGTCAATC AATTCAGCCG CGACTTCTGG GCGCAAAATT AGCCCGCTGC CTTTTCCACG 3540  
 CTCAGCCACG TTAGGCACAG GACAGCCCAT ATTTATGTCT AACCTTGAA AGCCCATCTC 3600  
 CGcTACGCCG ATACTCATTT CACGGAAAAA TTCCGGTTTA TCCCCCAAA TATGTGCCAC 3660  
 CATCGGCTGT TCATCTTCTG TAAAAACGAG GCGCCCGCGT AACTATCTT TTCCTTCAGG 3720  
 ATGACAATAA CTATCCGAAT TAGTAAATTC TGTAATAAAC ACATCTGGTG CCCCAGCTTC 3780  
 TTTACGACA TGACGAAAGA CCACATCAGT GACATCTTCC ATCGGTGCTA AAACAAAAAA 3840  
 GGGCTTTGGC AATGTTGCC AAAAATCTT GTCATCTTC ATATACTCCG TTTCTGATTT 3900  
 TTTCyTTAWT TTCAACGACA AWTGTCGACA TAATCTGGGG CGAAACAATG CTAATT 3956

## (2) INFORMATION FOR SEQ ID NO: 78:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7053 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

AAnCAATTTT TGgATTAgCy GCTGGGTTTT CCAAnGnGTT TTTAgCkGCT AAAAGCGCTT 60  
 TTgGTAAcTC TGTAACGAGG AGGCAGTAAA nTGGTGCCct TCACTTGGAG TGAACCTNyG 120

CGCTCGTtTG ACTACTTCTT GTAATTGGTC ITTATTTACA AGGGGTACTT CTTTTAATTG	180
AGCCATTGCA GTTTCCAATG CGTTTAGAGC CTCTGAGACT TCTTCTTTTG TCGCCGTtTC	240
TTTCGCTACG ATTGCTTCAG CCGCTTGAAT TGCTTCTTGC AATTGTTTTT CTGTTTCTTT	300
TGTAAATTGG TATCCTTCTT CCGGCTTTCT TCCTCGGGCT TGAGCGATTT GCTCTTTTAA	360
GGTTTCCTTA TCAACGGCGA CGGGTTGTTC TTTTAATTGT GCAAtTGcTT CACGCAATGT	420
TTTAGTCGCT TGGTCTATTT CTTCTTGAGT AACTTTTTCT TGAGCAAATA CTTTTTGAGC	480
TGCTGCTAAT GACTCTTCGA GAACTTTTTT AGTTTCTTCG GTGAATTCCT GCCCATCACT	540
TGGTTTCTTC GCTAGCCCCT CTTGAATTGC TTTTTCTAAA TCAGCTTTAT TCGCTACATC	600
AGGCATTTTT TCCAACGTAC TTGCAGCCTT TTCTAGGTTT TCCGCGGCAG CAGAAACGTC	660
TGTCTGTGTG GCTTGTTCAT TGGCTAGAAC AGCTACTGCC TCATCTAAGA TTGGTTTAAA	720
GATTGCCCAA GTTTCAGGCG TATATTCTTT TTCGACACGT GCTTGATTTA GCGTAATTAG	780
CGCTTGTAAT AACTCTTTAT CTACTAGCTC TGTGGGCTTC ACTGTTACAG AAAAATCTGA	840
AAGTGTGGCT TGTtGGAATC CTTGGCCTAC TAATAAGCCT TGACTCATAT CAAATTGTAA	900
TGGCTGCTCA ACAGTTGTCG CACGTTTTAC GCGAGTTGTT TCTTTTACTT TAACTTTAAA	960
GGTAACTAAG TCCATTGAAC CATTTAACAA TTGTTTATTT CCTTGATTAC TAAGAACTAA	1020
ATATAGATTC TCCACATCAT TTGAATGTTT CCGATATTTA GAATAATTTT CCATTTGTTG	1080
TGTATTATTC GTGGTTGCAG GCTGACCGAC AAGTTCAAAT TTGCTAGAAT CAAAGGACAT	1140
TTTTTGCTT AACGCATTAA TGTTTTTTAA GTCTTGACCT TTAAGTATAA AGGTTAATGT	1200
ATCGCCAGGA AGATAAGTAT CCTTGCCATT TTCATTAAAC ACAGCAAGGG ATAAACCGCC	1260
CGCAATTCT TCGACATCAG GTATCTCAAT CCCACCATCT AATTGTCGTA AAACATAACT	1320
GATATCATAA GCATCAATTA CCCATTTTTT ATTTAAATCA CCTTTTTCAA CATAGCCATT	1380
AAAGTCACTG TCCACAGATT CTAAGCCTGT ATAGTTTCGA TAACTCATTG CATCATTTTC	1440
ATCAATCGTT CCATCGTTAG TAATGTCTCC ATGTAAGATT CCTTCTGTTC CTGGTTGTTT	1500
GTAAAAGAGC ATCTCACGAC CAGAGCCAAA ATTACCCACC GaTTTTAATA CCTTCATTTc	1560
AACAAAACGa TACGCTTGGt CCTTCGTTTc AaKGGTTTTT GtCAAAGCAT CTTGcTTCCA	1620
GTTAATAGGT TCTGAAAACt CTGTCCAATT AGCTCCATCT TTCGACGTCC GATACTGGAG	1680
CTGCAAAATA TTTCCGTTGC CGGcATTGTC TCTTGCAAG TATTCAATTT TATCCATTTG	1740
ATACTCCGCA CCAAGATCAA ATTTTAAAGA CAAGAAATTT CCATCACTTG GATTAGCGAT	1800
GCCTGTTGAC CAGTTCGTAT GCCAACCTGT ACTCAAATCT TTATCTGTTA ATTCTTCAA	1860
TTCTGCGCCT GGTtGTTCTG GTAAATTACT TGTTGCTTTG ACTTGATTAA TTGTTTCTTT	1920
ATAGGGATCA TCTTGTGTTT TTCCTTTAAT CGGTTCACTC CATTcAGAGA CCCCgTTTTT	1980
TCCAACAGCT CGAACTCGAA AAGTATGTTT GCTTAGGAAG GAAAAATCCAT CAAATGTTGC	2040
TGTATTcGTT TGAATATTTT CGrAAACAGT CCCGTcACGC TCTACTTCAT AGCTTGTGTC	2100

CTCAGTTACT	TGATCCCATT	GAAGAGTCAA	AGAAGATGGC	GTTGTTTTTT	CTTCATTGAT	2160
TGCGACATTT	GCCGGAATAG	TTAACTGGTC	ATCGACCGTT	GTATTTCTC	CGTCAACCGT	2220
TCCTTTATTA	ATGAAACCTT	TGACAGTTAT	TTGAACATCT	TTGGCGGTCA	CAGATTGTTT	2280
CGCTAGTTTT	ACTGATAACG	CTGATTGGTT	CAGTTTTTCT	CCTGAAGCTT	CAGAAAAGATA	2340
CTGATTCACT	TGGAATTCTT	TATCAAAATA	ATATAGATTT	GTGCCAGCTA	AAAATTCTTC	2400
TTTATTAGCT	GCTtGTTTCA	AGGTTACTTC	TGTCCCCCCC	ACCATTGCTG	TTACGCTTTC	2460
GGsCGCTtCT	GAAGCTAGCA	AwTCAAGCGt	GGTGCTtCGT	tCATCAACAA	AGnCTTtATA	2520
ACTTCCTTTt	GTCGGTtCGA	tAGTaACGGT	CAGATCCCTt	TTcGTTTGAT	TTAGGGCCTT	2580
GACTATTTAT	TTTAGTCGTA	GCTGATTGCC	CAGCTTCATA	ACTAGTTGAA	ATTCCGTCAT	2640
CTTCATACAT	ATTAAAAGAA	GTGGTGCCAT	TTGGATAAAT	CAAGAAAGAT	CGTTGATCTC	2700
GTTGGATTTT	CTTTGGATTG	TTGTTTGGAT	TCGTCATTGG	GATAATACTG	CCATCTTTGA	2760
CGAAAACGGG	CACCTTCCAC	AGCGGTGTTT	TTACTCCATT	TAAGACACGG	CCACCTTGAT	2820
ATTTTTCGCC	TGTAAATAAG	TCAACCCAGA	CTTGTTTTTC	ATCTGGTAAG	TAGATTCCAT	2880
CACGGATTGA	ATTTCTGCC	TCATCTTGAT	TGCCATTATA	AATTGGGGCA	ACGAGTAAAT	2940
TTGGTCCCCA	CATATATTGG	TATTGAGAAT	CTTTTGATA	TGCTGTTCTT	TCATTAGGAA	3000
ATTCTAAAGC	CATGGCCCGT	ACCATTGGTA	GTCCATCAAC	TGATTCCTTG	GCAATACTGT	3060
AGTTATAAGG	CATCATCATT	GATTTTAATT	TTAGATACGC	TCGGTTCAAA	TCTGTGGCTT	3120
CTTGATCAAA	AGCAAATGGC	GTTTTTGGAT	TGGAGCCCCA	GCCGTCCATA	TTTAGCTGCA	3180
CTGGTGTAAG	AGTTTTCCAT	TGGAAATCAC	GTATGTTGAT	CTCTTTATTT	TTTCCACCAA	3240
AAATTCCATC	CATATCTGAA	CCGACATTTG	GTTGTCCAGA	CAACGAAGTC	CCGATATATG	3300
TTGGGATATG	GAAGCGAATA	TATTTCCATT	GACCACCCGT	TTGATCGCCT	GTCCAAATAC	3360
CAGCGTGCCG	TTGGGTTCCG	GCCCAACCAT	CTAATGAAAC	AATCATCGGA	CGAACTGCGC	3420
CGTCTGTTTC	TTTTACAAAA	ACATTGCGCG	CATCTTCTAC	ACCATTTAAG	CCAAAGGAAT	3480
ACCCATACCC	GACCCACGCC	ACGTCAGTTT	TGAGAGCTTT	AACACCAGCG	ACACTAACTT	3540
CTTTTCGCAAT	ATCCCGTTCT	CCTTTTTTAG	GATTTTTAGG	ATCCGCTGGA	TGGAGATTAG	3600
ATTGAGTCCA	CAGCCCCACT	TCAACACCAT	TGGCTTGAGC	ATACTCTGTA	AATTCTTTTA	3660
AGTTTTGTAC	ATCGCCGTCC	AATGAATCTG	TTTGCCATA	GCCAGCACCG	TAGCCGTCAT	3720
TTGGCAAGAA	CCACCCTAAC	GGCATATCGT	TTTTCTTATA	TCGATCAATA	ACTGCGCGTG	3780
CAGAAAATTG	ATAATTCTCT	TTTTCTCCAT	TTAAAGATTC	TAGCGTGCCA	TTTAAGTTTC	3840
CTAGATCGCC	TGGTTGGTAT	TCTTTGTAAG	AATTACCATC	TTCAAACCTA	ACTGCTCCCG	3900
CGGTTCTTTC	TGCTACTTTC	ACCCAATAGT	CCCGATTATA	CGCGTTTAAA	TGCGCTTCAT	3960
AGAACCCGTA	TTCTGGCATC	AAAGCTGGTT	TCCCAGTTAA	TTCATAGTAA	TCTTTTAAAA	4020
TACCTGCTGA	GGAATCATTa	AAGAAATAAA	AAGCATCAAA	ATCTGTTCct	TCGTGCGTTG	4080

TTGTGGTTTT TTGTGGATCA TGACTTCCAA ATCATAGTT GCCAGGTTTC CATGTATTTT	4140
GAACAACGCC ATACCCGCA GTTGACCAAT AAAATGGGTT AGCGATGCT ACTCCGCCAT	4200
CTACCCAAAT GTTCGTATTG ACAATCTGGA TAGCTGTCC TTTATGCGTA AATCGACCAT	4260
TTTGAGTTCC TCCGCCAAA TAATTTTCTT GACTACTTG TTTAACGTT TGTGTAGCTT	4320
TATCATTTTT AAAGGATAGC GGAGCGGTTT CTCTAAAT AACTTGATTC TTTTCTTGT	4380
CTAGAACCTT CATTAAAGCG CTTTCTTTT CSAACATAAT TTTTAAACC TTTTCTTCTA	4440
GAATAAATTG GTTTCCTGAA TCTGTGACGT TCGTTTGCTC AAAGGCTTGT GTACCATAGT	4500
CGGCCATTGT TTAGCCGTA ATTCTCGCA CTGCTTIGG ATCATTGGC GTTGGATATT	4560
CTTCAAATTT ACCAGTGGGA TCTAGATGAT AGCGAAAAG ATGGTCGTTT AAAATACTAA	4620
TTTGTGCTTT TTTCCAGAA GCATACGTAA TGACAAAGTT TTTCTCCGCT TTTtCAATCG	4680
ACACAATTGC TCCGACTGAC GTTGCTGAAT CAGCTACGTC TTCTTTTGCT GTTTGCTCTT	4740
GCGCATCAAC CACTTGAGAA TCCATCGTAA CAAACGAATG GAACGCCAAT ACAGACGAAC	4800
TCACTAATAA AATGTCAAC CAGTGCTTTT TACCGTAAAT TTTTACCC TGATACTCTT	4860
TGTTCATTTT TTCTCCTTA AATTGCTCC CCATAGAATA GTTCCGCAA AATTACTCAC	4920
AAAATCCAAA ACCCAACAAA GTTTATCCA GTGAAAAGT CATAACGATT TTCTTAGTTT	4980
ACTGAGACTG TTAGAAAGTC AGTTTCAGTA ACCCCTTCAT AAACACATCC tTTCAATAAA	5040
ATATTAAATG AATCAATCT ATTTTCATCT TATATAAAA TTTTAGTACA GAATCATAAT	5100
TTCTTAATGA ATCGTETTC ATTTGTTGAC TACCAATTT TCACTTTTTT AATAACATAA	5160
TTATCGTATG ATGGTTTTT TGTGTTTGT AACGGTTAAT TATAGCAGCC ACCATTAACA	5220
AGTATTAACC CTGATAAAC CTGTTAATTC ACTTATAATG ATGATTTTAT TGCGAAAATA	5280
TCTACATCAG AAAAAGAAAG ACTTTTTCAA AATCTAATA CCTGTTGCTC AAATCTAGTC	5340
TATTTAAFAA AATCAACAC CCTCTGCTC ASCAAGAGGG TGTTAATTAC TCGTTTTTTA	5400
GTGATTGATT TGAATTGCTG GTGTGAAAAG CAACTAATG ATGATGCCAA CTATCAGTAC	5460
CGCTAGCAAT ATAAATTTA ACTTTCCAAA CCGGTGATAG GGATTGATCG AAAGCCCAAT	5520
CCCATAACT TTCGGAACCC ACAAGGAGG CGTTGGATTA TCTTTTGTAT CTTTATAGGC	5580
TATCTTTAAA ATCTGCCACG CAAATACAA CATGGTAAA ACCATAATTA CCCCACTAT	5640
ACCTAGCCAT TTGCTAATA ACTCACTAGT AAGCAAAGA ATGATTGTCA GAATTGGTAC	5700
AAACAAAGA ATAAAGCTAA TGTGGCTCC CATTGTGCT TGCTTAAGCA GTCGTCTCTC	5760
ACGGAGAGTA AGATTGCAA AATTCCATC CTCCTTTCA TGGTTTCAG ATTAAATCAA	5820
CGAATCACTA AAAAATAAA TTAATGCGA AATCAATGA CCATGCCCCC GCGAATTTTT	5880
CCTGTTCCTA TTCTTCAAA AATAGCATT ATTCTCTG GTTGACSTGT TTGACAAACA	5940
GGAACAACCTT TTCCTTCTGC ACCGAATTGG AAGCTTCTC TTAAATCTTG ACGGGTTCCA	6000
ACTAACGATC GACTAGGTG ATGCCATCA AGAACAAGAC GTGGAATAGA TAAATCCATT	6060

GTCTCAACAG GCAATCCAAC GGCACAAAC	GTTCTCCCG CACGGACTGC ATCGACTGCG	6120
CCATTAAATG CTGATTTAGC TACCGCAGTG	ACTACTGCTG CATGGACACC GCCAATTTCT	6180
TTTTGAAGTT GTGCAGCAA ATCAGGATCT	TTAACTGGAT TAATTAAAAG ATCAGCACCC	6240
ATTTCTTTTCG CAAAAGCTAA CTGATTATCA	TTCACATCAA CAGCTACTAC TTTTGCATTA	6300
AAGACATGCT TGGCATATTG TAAAGCTAGA	TTACCTAGCC CACCTAATCC TGAAATCATT	6360
AGCCATTGAC CTGGACGAAT CCCAGATTCT	TTTATCGCTT TGTAGGTTGT TACCCCTGCA	6420
CAAGTAATAG AGGAAGCTGG CTCTGATGCT	AAATTTTCTG GCACTTTTAC TGcATAACCG	6480
TCTGTGACGA TGCATTCTTC GGCCATGGCA	CCATCAACTG TATATCCAGC ATTTTTTACT	6540
GAACGACAAA GCGTTTCTCG ACCAGAAGTA	CAATATTCGC AATGCCCGCA GCCTTCATAA	6600
AACCATGCAA CAGATGCACG ATCACCAATT	TTTAACTAG TAACGCTTTT TCCAACCTCC	6660
GTCACAATCC CAATTCCTTC GTGACCTAAG	GTACAACCGC TAACATCGCC AAAATCTGCA	6720
TTCTTAACGT GTAAATCCGT ATGACAGACA	CCACAGACTT CCATTTTCAA CCGCGCTTCA	6780
TTATCTTTCA ATGGACGCGA TTCTTTTTCA	ACAATTTTGT TTGTTTTTCC TTTTGCTGCG	6840
ACAATTGCTT TATACATGGT TCTTCTCTCC	TTGTTTCTAT TGTGAATCAT TGAACAATAA	6900
TAGTATAATA CTACTTTCCA AAATAAGGrA	CACCAATGAC TtGCGTAAAT TAACGCTAAT	6960
TTGaCAATTT ATCCTkGCTA GACAGCTAtc	TTCCCTTATG TCATAGATAA TTCATACTGG	7020
CATGAnTTTC TCTGCATAAT TGnCAACTAT	TTT	7053

## (2) INFORMATION FOR SEQ ID NO: 79:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2213 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ACTTATTTTA CTAACGTCGT GACGCTAAGT	TTAGGTGGAC TTGTAGCGAT AGTGATGTAT	60
ATCGATCGAA TTTATACGCC GATAGCCATA	TTTAATGTTT TATTTGTTCA ATATAATTTG	120
GnATAAAGTA GCCTATCAGC GCTTGGAAGA	TTTTTATAAA AAAGAAGATG ATCCTGACTT	180
AGCGGTTTCC GGCAAAGCGT TACCTGAAAT	TCAAACAATT TCATTAAAAG ATGTGTGTTT	240
TAGCGTAGAT AGTCAAACGA TTATGTCCCA	ACAAAATCGA CAATTTTCAA TGAATAAAAC	300
CTATGGCTTG ATTGGAAAAA GTGGGACAGG	TAAATCTACC CTGATTAAAT TGATTTTAGG	360
ATTATTGAAG CCGACAGAGG GGACAGTTTA	TGTGAATCAA TTTCCTCTAA CACAATTTAA	420
TTTGGAAGAT TATTATGAAA AAGTTTTTTA	TTTATCTCAA GATGTGCCAA TATTTCAAGG	480
AACGTAAAAA GAAAATATTG TATTTAATCA	AGAAATTTCT GATGAACAAG TAATTGAAGC	540
TATGTATCGT TTTCAGTTGG GGGAACTTTA	TGAACGGTTG CCGGAAGtTT AAACACTATt	600

559

GTGAGTGA AAA AAGGAtGAAT TTTTCTGGTG GTGAAAAACA ACGGATTGCT TcACACGCTT 660  
 AGCATTTACA CAAGCAGAGA TCCTTATTTT AGATGAGGCG ACTTCTGCaT TGGACGAAAA 720  
 GnCCGAAGAA AAAGTGTAC AAGAAGTACA AAAATTTACC CACAATAAGT TGACTATTTT 780  
 GGTTACTCAT CGACCTAAAA CCCTGCGATT TGTAAGTGAA ATTATTGATT TAAATGAGTG 840  
 AACTTGTTAG CTCTTTGGCG CTTCTTTTGC TACACTTTTT ATAAGAATTG AGAGAAAGGG 900  
 GCAACTGTTG ATGAATAATA ATGAGCAAGA ATTAGCGGAA TTTTGGGATG AATTTCGAGA 960  
 AGAATATGAA GAAATTCAAC AAGAATCGCC ATTTCCAATT GCTCGAGAAT TACGTGATTT 1020  
 CTTGGTCCAA GAAGGCATTT TTCCGTGTCA GACATTTCTG GATATCGCAG GTGGGACGGG 1080  
 ACGATACTTA CCTTTTTTTC AAGAGCAAGT CACGGAGTAT ACGCTAGCCG ATATTTCTCA 1140  
 GAGAATGTTG GAGATTGCGG AAGCAAAAGC ACAAAGCAAC GTTGTTTTTC TTCACCAAAG 1200  
 CCAAGAAAGA CTGATTGAGA CTGGGAAAAA ATTTCAAGTC GTTTTTTCAG CGATGAACCC 1260  
 TGCTTTGGAT ACGCCGGAAA AAGTTAATGC ATTGTGCCAA TTAAGTGAAG AATGGTGTCT 1320  
 TATATTTTCG CTAGTGGAAG AGCAAGATTC GTTGTTTTCT CCTTTTGAGC AAGAGAGCAA 1380  
 CCCGCAGTTA AAATGGATGG CACAATACAA AGCATTTCTA AAAAAAGAAC AGCGACCTTT 1440  
 TTTACAAAAA AAATTTTTCT TTGAAGCATC AGAAGCTATT TCAAAAGACT TTTCCGCAG 1500  
 CTATTTTGAA GAACAATGGT CAGTTCCAAT TTTAGAACAA CGCGTACAAG AAATTTTGG 1560  
 TTCTCATGAA ATTAAGcAAA ATCAGCGTAC CATTATTTAT GAATTAATTG CGATTCCATG 1620  
 TAAAAAACA ACCAGCGATG ACTAATTATC GCTGGTTGTT TTTCGGTTAT TTAATTGTGC 1680  
 GAAAGCTTTT TCTGGAATCT CTTTTTATC AATTTTGACA GGAGATTGGA TGATTCTTTT 1740  
 TTTACTAATG GTTGCTTCTA TGTAAGCATT TGGTTCCAAA GGTGAGGAT TTTTGTCTGT 1800  
 CCGTTTCGTAG TCCATAACTT GCGTTTTTCC ATTACTTTTA ACAAAGTTA GTCGATAGTG 1860  
 TAAAGAATAT TGGCCATCTA TGCTTTTTTG ATTGTCATCT ACTGTTTGTT TTCGTTCAGG 1920  
 AATTTCTAGA GGAACTTGTG CATAGGCTTT TTCACCCTTA TATGTATCTG CATAATAACG 1980  
 ATAGCCTGTA TAACCACCAA CAATGAATAA GCCAATAAAT AAAATAGCGC CGATTTTTTT 2040  
 CATTTTTTCT CCTCCTAACA ATTTACAAAA ATAGCATAGC AAAGAAGAA GGAGATGGAA 2100  
 CATCGAAAAA ACGAACAGTT ATCTTACAAA ACTGTTCGTT TTTTTGGTC TATTTAGCTT 2160  
 TCATCAGAGA GTAGAAAGGA ATCATTTTCT TACGATAACT AAAGAAAATA ACn 2213

## (2) INFORMATION FOR SEQ ID NO: 80:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4808 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

ACGTTTCAGGA ACTGATACCC TAGATTTGCG CAAGTTGCAT CGTCCGCAaA GATATCaTCA 60

ATCACCACCA	ACGCCGTGGC	ATTATTTTGA	TGAAGAATGT	CATCAAAAGC	CATGTAATCT	120
TCAGGTTTTT	CTAACGCAAC	GCTTGGTAAA	ACAGTTAATC	CTTGCATCAT	CATTGCTTTT	180
TGATAACCAA	AGTAACGTTT	AAAGTACAAA	CTTTCATGCG	TTGTATTGGT	CACAAATAAA	240
ATATTTTTAT	GGCCGTTTTT	AATTAAAAAT	TCCGTCGCTT	GTTTCCCTAA	TAATTGaTTA	300
TCATTATCCA	CGTAAACTAT	TTCCGTTTCA	TGTTGGTAAG	GTTGACCGAT	TAAAGTAAAG	360
GGAATGTTGT	TTTCGTCTAA	GTAATCAATG	ACGGGGTCCT	TACTATCTGA	ATACACAAGA	420
ATAAAACCGT	CCACTTGTTT	TTGAAGATGC	ATTCTTTTAA	CGTTTTCCAA	TAGTACATCA	480
AAACTTTTGG	CAGTTGCAAT	AGCAGTAGTC	ATATCGTAAA	GACGGGCTTC	CTCGTTGATG	540
GCTTCCATAA	TTTCTAAATA	AAATGGATTC	CCTAACCGTT	CTTTCGAATC	TAAAGGAGGT	600
AAAATTACGC	CAATGGCACT	TGAAATTCGT	TTACCCAAAT	TCCGTGCCGT	CATATTTGGC	660
ACATAGCCTA	AGTCATCCAT	GACTTTTCGA	ACTTTCTTTT	TTGTCGATTC	AGAAATACTT	720
GGATGATCGT	TAATGACACG	CGAAACGGTT	GAAGTTGCAA	CCCCGCTTT	CTTAGCGACA	780
TCTTTTACTG	TAATTGCCAT	TTTTGCCACG	CCTTTCTTAG	TAATTATAGA	AGGTTAAATT	840
GATAAACCGT	TTGGTAGGTT	TGCTTTTTAT	TAGCTGGTAA	TAAAATAGAG	CCCCACTGAG	900
GATGATGAAC	AATGTCTGGA	TACTCTTGTT	GTTCAATTGC	TAAGCCATAG	TTTGAATGCA	960
TCGCTTGTC	ATTAATAGAA	AATGGAGCCT	CAAAACCACT	CGTAGAAAAT	AATACCATAC	1020
TTTGGTTCGT	TGTTGCTATC	GTCATTTTTC	TTCCAGAAGT	AGGCTCTTGT	AAGCATAATT	1080
GTGGCTGTCT	TAAGTCAGAG	CCAACTAAGG	AGAAAACATC	ATCTAACCCA	GTTGGATAGC	1140
GTTGGAGGAT	TGTTTCGATA	GGTGTCAGTT	CATTAAAACG	TATGGATAGT	TCAGAAGCTG	1200
GTAAAATCCG	TCCAGTAGGC	AGCTTATTAT	CGTCTAATTC	TAACATACCT	TGACAGTTCA	1260
AAGATAGCAA	ATGTGACAGA	ATGTCTAACT	TTCCATCGCC	AGATAAATTA	AAGTAGGCGT	1320
GGTTCGTAGG	ATTAAAAAGT	GTTTCCTGTT	TACTCCGTCC	AGTACTGATC	aTTCGTAAAT	1380
GATTATCCTC	TAGCTGATAA	GTGATGGTGG	CCTCAATTGG	TCCAGGATAG	CCAGAAATGG	1440
TGTCGGTTAA	ATGAAAAGCA	ACGCCAATGA	TATCCTTTTT	TTCAAAGGGC	TCGACTGTCC	1500
AATATTGAAA	AGACCACCCC	TTTGAGCCAC	CATGGATATG	ATGACCGTCG	CTGTTTTGTT	1560
CTAATATTCC	ATTTTGCCAA	TGGCCATTAC	GTATGCGTCC	AGCAACTGGT	CCAACAGTGG	1620
CACCAAAAAA	GCTTTGATCC	TTTAATACAT	CTTGATACGA	ATCATACGAT	AAAAGGATAT	1680
TTTCCCAATG	ATTATTTTTG	TTTGGCATCA	AAAGTTGATG	GAGGCGCGCG	CCATAATTGA	1740
GAAGAACGAC	TTCTAAAAAG	TCGTTCTTCA	AAGTAATTTT	TTTAAAGTTT	GTGTTTGGA	1800
TATCGGTTAC	GGTGATACGA	TTATCCTTCA	TGATCAAGCC	ACACTTTTTT	TAAGTAATCA	1860
AACGTTAAAG	CGCTCGTCTC	TGGCACGATA	GGCAACCCTT	CGCCTAACTC	TTCTTTGCGG	1920
CCAACGCCAA	CAGGTTGCGC	TCCACTAGCA	AGAATCGCTT	GAATCCCAGC	TTTGGCATCT	1980
TCAATGCCAA	GACATTCTGC	TGGTGTCAAA	TCAATTTCTT	TCGCAGCTAA	AATGAAAATA	2040



TCTGGAGCGG	GTTTACCTGC	CTGGACGTCT	GCTGGATTAG	CAATGGCATC	AAAAAGGGGC	2100
GTCAATTGCA	TTTTTCTAA	AAGAAAAGGC	CCATTTTAC	TTGCAGAAGC	CAAGGCGATT	2160
TTAATATTGG	CTTCACGTAA	TTCCGTTAAT	AAAGAAAGAA	TTCCTGGGTA	GACATCTTCA	2220
GGCGTAATCG	CTTGGATCAT	TTCCAAGTAA	TAATCGTTTT	TCCGTTGCGC	TAAGTGGCA	2280
AATTCTTCAG	AAGAAAACGT	GCCTTCTTTT	TTTCCGTATT	TTAACAGTAA	CTGCAGAGAA	2340
TCTTCACGGC	TTACACCTTT	CAGTTGTTCG	TTGAAAACAC	GGTCAATGGA	AATTCCAATT	2400
TCGTTCCCTA	ATTTTTTCCA	AGCGTGGTAA	TGAAATTCTG	CTGTATCAGT	AATAACGCCA	2460
TCTAAATCAA	ACAAAACGCC	TTTAAACATG	AACTTCTTCC	TTTCTGATCG	GTAAAACTAA	2520
TCGATCTTTT	AACGTGTATT	TTTGATCGTA	AACAGTTAAT	GGAATTGCAG	GACCGTCGAG	2580
TAAAGTGAAG	GCCACTTGGT	CTGCTGCAAC	TTCAACAAAT	AGTAAACGAT	TCCGATAATT	2640
AATGTGGAAT	GAGTAGGCTG	TCCATGTCGC	TGGTAAAAAG	GGTGCAAATT	TTAGTTGTTG	2700
ATGGTCCGTT	TTCATTTGGG	CAAAACCTTG	AACAATCGCT	AGCCAACTAC	CAGTCATTGA	2760
GGTAATGTGT	AAGCCATCTT	CTGTATCATT	ATTGTAATTA	TCTAAGTCAA	GGCGCGCTGT	2820
TCGTTGATAC	ATTTCCACTG	CTTTTTCTTC	CATGCCTAAT	TCAGCAGCTA	ACACAGCATG	2880
AATGCTTGGT	GAAAGAGAAG	ATTCATGGAC	AGTCATTGGT	TCGTAAAAGT	TAAAGTTCCG	2940
GCGTTTTTCT	TCTAAAGAAA	AGGCATCATT	GAAGAAATAA	ATTCCTTGAA	GAACATCGGC	3000
TTGTTTAATG	AAACAAGACC	GTAAAATTTT	ATCCCATGAC	CAATGTTGAT	TTAAAGGCAG	3060
CTCACTAAGT	GGTAAATCAG	AAACAGGCAT	TAAATCTTTA	TCTAAGAAAG	TGTCATGTTG	3120
TACAAAAATG	CCTAATTCGT	TATCCACTGG	GAAGTACATA	TTTCTACAA	TGTCTGCCCA	3180
TTTAGCTAAT	TCGTCGCTCG	CAATTTTGAG	CGTTGTTTCT	TCTTGGAATT	TTAAATAAGA	3240
CTCACGGGTG	TAACGTAGAA	CCCACGCAGC	GATGGTATTG	GTATACCAGT	TATTATTGAT	3300
ATTATTTTCA	TATTCATTTG	GCCCCGTTAC	GCCATGAATC	ATATATTGTT	TGTGACGTTG	3360
GGAAAAGTGG	ACCCGATCTG	CCCAAAAACG	AGCGATTGCT	ACAAGTACTT	CTAGACCAGC	3420
ATCTTTTAGA	TAATCTTCGT	CTCCTGTATA	GTTGACATAA	TTATAAATCG	CATAAGCAAT	3480
CGCACCATTT	CTGTGAATCT	CTTCAAAAGT	GATTTCCAC	TCATTATGAC	ATTCGACACC	3540
AGTAAAAGTA	ACCATTGGGT	AAAGCGCACC	CTTTAACCTT	TGTTGTTGGG	CATTATGAAT	3600
GGCTTGAGGC	AATTGATTGT	GGCGATATTT	CAGTAGATTT	TTAGTAACTT	CGGGCTTAGC	3660
CAACGCCAAG	TAAAGTGGGA	CAGCATATGC	TTCTGTATCC	CAATATGTGG	CACCTCCATA	3720
TTTTTCGCCC	GTAAAGCCTT	TTGGTCCGAT	GTTTAAACGA	TCGTCTTCTC	CATAGTAGGT	3780
TGAGAAAAGT	TGGAACAGAT	TAAAACGAAT	GCCTTGTTGG	GCTTCGTCAT	CTCCTTCAAT	3840
AACAACATCA	GCTAATTGCC	AACGTTTTGC	CCAAGCGGCA	GCTTGCCCAG	CTTTTGCTTC	3900
AGGGTACAAG	GTAGTCATTT	CTGCAAATAA	TTCGTTTACT	CGTGTGATTT	GTTGCGTTTC	3960
TGGCACATCA	CGACTGGTAA	CGACAAGCAC	TTCTTTTGTG	ATGGCGGTCTG	TTtCGCctTC	4020

GTTTAAAGAC aCTGTCAATT CTGCGGTAA AGCCAAGGGC TGTTGTGTAT ATGTAGGCAC	4080
AACAGAGGCT CCATCGATGA AATGGCGCAT GCCTGCAGTC ACTGTAAAAC GCTCTATTTT	4140
AAAGTTGTTA GGAATGGTCT TAGTCGTAAC AAAGCCAAGT GTTTCTTGTC TTTCAATAGC	4200
TCTTTCTTCC CAAAAATGTT CTTCATAATT GCTGTCTTCA TTTTGAACAT CGCCATCCAA	4260
TTTTGAGTGA AGAGTGATG TGCCTGTACC TTCCAGCATT TCAATTGTTA AGTGGATGTA	4320
AGCTGCCTCT TTTTTTCTA AGCTTAAAAA ACGTTCAAAA GAGAAGCGCA CTTTATTCTT	4380
CGGTGTTTGG ATGGTGAATT GCCGAGAAAG AACCCCGTTT TGCATGTCTA ATTCTAAAGA	4440
AAAATCTTCG TAAGGTGTAG TGGCTAAATC AATTGTTTGT CCATCAATTT GTAAGTCCAT	4500
TGCGATGAAA TTAATTGCGT TAATCACTTT ACCAAAATAT TCAGGATAAC CATTTTTCCA	4560
CCAACCAACA CGTGTGTTTGT CTGGATACCA AACACCTGCT AAATAGGTTT CTTGATGGTG	4620
GTGCCCAGAA TAGTGTTCTT CAAAATTTCC CCGCATCCCC ATATAGCCGT TGCCAATGCT	4680
TGTTAATGAT TCTTGTAAC GTAAATTTTCT TTTATCTAAG TGTGTAGTAC GGATTTTCCA	4740
AGGATCAATT TGAATAAAC GTTTGATTG TTTTCATGAGA ATCCCCCTCA GTGATTTTAT	4800
TGATAGGT	4808

## (2) INFORMATION FOR SEQ ID NO: 81:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15747 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

GAGTAGGAAA CAGATATACT TAATATAGAA AATTCAGATG CACTTATTGG GAGGTTATTT	60
TATGCAACGA AAAGATGTTA CACCTAATTC AGATAAAAAA AGTTTATTGC AAAAATTGG	120
GATTGGTTTA GCTGGCGGCT TACTTGGTGG CGCCCTTATT CTAGGAGGCG CTTATTCTGG	180
GATTATTCCC ACACCCAACG GCGGAAATAA TGCGGCGACG ACGACATCCA CTAATCACGG	240
CGACACAAAA GTCAGCAATG TAAGTTACAA TGTGTCTAGC GATGTCACAA AAGCCGTGAA	300
AAAAGTTCAA AATTCTGTGG TTTCTGTCAT TAATATGCAA AGTGCTAGCA ACAATTCTTC	360
GGCAGATGAT CCTTTTGGGG GATTGTTCGG TGGGAATGAA GGTACTCAAG ACTCTTCCGG	420
CAACAATGGT AACGATTTAG AAGCAGCCAG TGAAGGTAGT GGGGTTATTT ATAAGAAAGA	480
TGGTAAAACG GCTTATGTCG TGACAAATAA CCACGTCGTA GATAAAGCGC AAGGATTGGA	540
AGTTGTTTTA TCTGATGGTA CCAAAGTGAA AGGCGAATTA GTCGGAACCG ATGCTTACAC	600
GGATTTAGCC GTGATTAAAA TTTCTTCCGA CAAAGTTGAT CAAGTCGCTG AGTTTGGA	660
TTCTAGTAAA ATCACAGTCG GTGAGCCTGC TATTGCAATT GGTTCCCCTC TAGGCTCTGA	720
TTATGCTAAC TCTGTCACAC AAGGAATCAT CTCTTCTGTG AATAGAAATA TTACCAATAA	780

AAACGAGTCT	GGTGAAcCAT	CAATATTAAT	GCTATTCAAA	CCGATGCTGC	CATCAATCCA	840
GGAAACTCTG	GTGGTCCACT	AATCAATATT	GAAGGACAAG	TCATTGGaAT	TAACTCAGTA	900
AAAATTGTGC	AATCAACTAG	TCAAGTGAGC	GTTGAAGGGA	TGGGCTTTGC	CATTCCAAGT	960
AATGACGTAG	tCAACATTAT	CAACCAATTA	GAAAAAGATG	GTAAAGTGAC	GCGGCCAGCT	1020
CTCGGGATTA	CGATGTCTGA	TTTAACAGGT	ATCTCTTCAC	AACAACAAGA	ACAAATTTTA	1080
AAAATTCCAA	CTTCTGTAAA	AACTGGCGTA	GTGGTTTCGTG	GTGTTGAAGC	AGCGACCCCT	1140
GCTGAAAAGG	CTGGATTGGA	AAAATACGAT	GTTATCACGA	AAGTTGACGG	CCAAGACGTA	1200
AGCTCTACTA	CAGATTTACA	AAGCGCGCTT	TACAAGAAAA	AAGTTGGCGA	CAAAATGGAA	1260
GTGACTTATT	ATCGTGGTTC	TAAAGAAATG	AAAGCAACTA	TCGATTTAAC	CATTGATAAA	1320
TCAGCTTTAA	CACAACAAAA	TAATCGCAGC	AATCAATAAA	AGAAAGAAaAC	GACCTGTTGT	1380
TCAAACAACA	GGTCGTTTCT	TTCTTTTATT	CTCCTACCGC	AAATGCAGAA	CCAACCTCTG	1440
CATCAAAAGG	CAATTCTAAG	ATTCCTTTTT	TAGCTGGTGC	ATTTGGTAAA	CGTAATTCTT	1500
TTGCTGAACA	AATCATTCCA	AAACTTTCTT	CACCACGTAA	AACCCCTGGC	CAAATCATT	1560
AACCATCAGG	CATCATTGCG	CCTGGTTTGG	CCACAACCTAC	TTTTTGTCTT	GCTTTGATGT	1620
TCGGTGCGCC	ACAAACAATT	TGTAAGATTT	CGCCGTTGTC	TACTTCTGTT	TGTGTGATAG	1680
ATAAATGATC	TGAATCTGGG	TGTTTTTTTAC	ATGATTTGAC	AAAGCCAACC	ACAATTTTAG	1740
GGGTGTCGTC	TTTTTCAAGG	GTTTCTGAAA	ATCCCGCATC	TTTAATCGCT	GTATTGACTT	1800
TTTCGATTG	TTCGTCTGTT	AAAGTCACTT	GGCCGGTGCC	ATTAATGTCA	CCTAGAAGCG	1860
CTGAAGCATT	GAAATATTC	CAAGCTACCA	CTGTTTCTGG	CGCCTCAACT	AAACTTACAC	1920
GTGCCACATC	GCCGACACGT	TTTACTTGGT	TTTCCGCCCC	TTGATCGTCC	GCAACGATGA	1980
CAAGTAAGAC	ATCGCCAACG	TGTTCTTTGT	TATATGCAAA	AATCATTTCT	GTTCCTCCTA	2040
AATTTGAAAG	CTATAGGCTG	GaTAGCTAAA	TTCTCTCCTG	TCCTATAGTA	ACAAAAAAAG	2100
TGGaAAGTTG	CACTTACAAT	TATTGAATTT	TAGGCAATCC	CTAATGCTGC	CAAAAAAATA	2160
GATAATAGAA	TAACCCCAAC	TAAAAGTGTC	GTTGTTTTCA	CTTTTTTCCC	TAATAACCAG	2220
TAAATAACTA	AGAAGAATAA	AGCCGGGAAA	AAACCTaACA	TGATTGGTGC	GATATATTCT	2280
TGTAATGGCG	TGGCGATTTT	TCCAGAGCCA	ACTTTGGCAG	TTAACTCAAA	CACGACATTA	2340
GAAGCCGTCA	TCCCACCAAT	AACCATTAAC	CCTAAAATAG	AGGCAGCTGT	GGTAACTTTT	2400
TCGATTACGC	CGCTATCTTC	AACATCACTA	AAAAATTTAG	CACCTAATTT	AAACCCTTTA	2460
TCCAAACAGA	TATACCGTAA	AGCAAAATGC	GGAATATTGA	TAATCAGTAA	AAAGGCAATC	2520
GGACCAAAAA	TATTTCTTGT	TTTAGACAAT	GAAATAGCTA	CACCTGTTGC	GATAACTTTT	2580
AAGGTCCCCC	AAAAGAAAGA	ATCCCCAATC	CCAGACATTG	GTCCCATTA	ACTAGTTTTC	2640
ACTGCATTAA	TACTAGCTGA	GTAAATTCA	TGATTTTTAG	CATTTTCTTC	TTCCATTGCT	2700
CCAGAAATTC	CTAATAATAA	GGTAGAGATA	TGCGGGGTGA	CAGACATAAA	CTCTAAATGT	2760

CGTTCTAAGG	CCGCTTCTT	CTCTTCATTA	TTTTTATACA	ACCTACGTAT	AATTGGAGCC	2820
ATTGCATAAG	CATAAGCTAA	ATTTTGCATC	CGCTCATAAT	TCCATGATGC	ATCTAACGTT	2880
AAAGAGCGAA	AAAATACAGA	CCAAAAATCT	TTTTTAGTAA	TGACAGATTT	TGTGTTTTTA	2940
GAATTCATTG	TCATCGACAA	CAGCCTCCCT	TGTTTGACCA	TCACTCATAT	TCATATATAG	3000
TAGGACAATC	ACGATGCCAA	ACAAAGCAAT	CCCCACTACT	GAAATTTTTA	AATAAGCAGC	3060
GAATAAAAAG	CCTAATAATA	GAAATGGAA	CaTCTTTTTT	GTTACCATCA	TTCTTGCTAA	3120
CATCGCAAAA	CCAATTGCTG	GaATaATACC	TGCCGCAATA	TCTAATCCaT	GCGTAATAAA	3180
GTCTGGGACA	TAAGACAATA	CCTTTTCAAT	AAACGGCAGC	CCTAAGTAAA	AGCCTAATGC	3240
CACAATTAGA	GCTAACAAAA	CTTTATTGCC	AACTCCTGCA	ATTAGTTGAA	TTCGCTCTAC	3300
CCCTTGAATA	TTGCCTCGTT	TTGCTTCCTT	ATCAGCTAAA	TGGCTAGCGC	CTGTAAGTAA	3360
AAATGTCATC	TGAAGATTAT	CGATTAAAAG	TACTAATGTA	GCGATTGGAA	CAGCTAAAAC	3420
AACCGCACTA	GAAATTGATT	GCCCACTCAT	AATTACATAA	GCTGTTGCAA	TAATCGTTCC	3480
TGAAACAAAA	TCTGGTGGAT	TACTTGCGCC	AACAGGAAAT	GCACCCATAA	ATACCAACTC	3540
TAATGTTGCG	CCAAGCTTAA	TTCCCGTATC	TAAGTCTCCT	AAAGCTAACC	CTACTAGCGG	3600
AGcGACAACA	ATTGGACGGT	TCCACATCGT	TGAACCAAAC	CAGCTATGCA	TGTAGGCAAA	3660
AAAGGCAATT	AAAAACATTA	ATAGAACCTT	GATAAATGTC	ACTTCCATCT	ATTTGACCCT	3720
CCTATTTATT	TAATTTAATA	ACTTGTTTAT	AGCAACCTTT	TGATCGTCAG	GGACTAATTG	3780
AACGGTTAAT	TCGATTCTT	CATTAGTTAG	TTCTTTAATT	TGTGCAATGT	CAGTTTCTGA	3840
TAAATGAACA	GCTTTTGAAA	TTTGTTTGCG	CGTTTCCCA	TTTTTCATTC	CACCTAAATT	3900
AACTTCCTTA	ATTGTTGGCA	CTTCTTTGAC	CAATTGATAC	ATATCAGCAA	CTGATTCACA	3960
TAAAATAAAT	AGTTTATATT	TATCTGTCAC	CCCAGAATTT	AACGCCTTAA	tTGAATCATC	4020
AATCGTTTTT	ATAACTAATT	TAACACCTGA	AGGTTTTGCC	ATTTTCATAG	CGCTCATTCT	4080
TAATGGATCC	TCAACCAACG	TATTACTAGC	AATCAAAATA	CAATCTGCTT	CTAAAAATTT	4140
TGTCCAAGTA	AACCCGACCT	GACCATGAAT	CAACCGATGG	TCCACTCTTG	TCATTTTAAT	4200
CATCCTTATT	CACCTCTCAG	TTCTTTCACA	ATTTTCATAA	ACGAATCAAC	ATAAGGAGCA	4260
TAAATATCTC	CCGTATCTTT	TCTTGATCT	TTGAAGTTAC	TCCCTACAAT	TGCTGCATCG	4320
CAAATCGCCA	ATTGTTCTTT	CACACTCTTG	TCATTTAACC	CCGCTGCGAC	AATTAATGGA	4380
AACTCAGGTA	ATTGTTTTCT	AAACAACCTA	ATCTTTTCTA	GTGATGTCTC	TTCACCCGTT	4440
GCATTTyCTG	TCACTGCAAT	AGCATCACAA	CGTTGCTGTG	CAATTTTCAA	ATCTTCTTCG	4500
ACTGATTTTT	CAGACAACAT	TGGTTGATAT	TTGAAGCGCA	CGCCACCAAT	TAATTTTGCT	4560
GTGTTTTTTG	CACGATATAA	ATCAAAAAAA	GCTTGTAAGG	ATGCTTCATC	CCGCGGTTTC	4620
ACATGACCAA	CCACAGAATC	AATCTGTAAA	AATTGTAAAT	GATATTTATT	GGCTAAATGA	4680
AAACCTAATG	GGTCAACATT	TAACACATTT	ACACCAATAG	GAATCGGTAA	ATCTAATGAA	4740

GTAACATATT GGaGCGCTTT TTCTAGCTGA ACATAATCCC CATAATAGTT CTCCATCAAA	4800
ATkGCATCTA CCCCATGTTT ACTGTAAATT tGAATCTCTT TTTTGTCTCG CTCTTGAATC	4860
TCTTGATCTG TTTTACCTTT TAAATGAATC ACCCCAATAA TTGGTTTTTC AACAGCAAAC	4920
AGCTCTAAAA AATCAGCTTT TTTCTTCATT AATACCCCTC CTAGAAATCA TCATTTTCTA	4980
TCTCTTCGAT ATGAACGCGT TGAATTCCTT GTCGTGCATT TCGAATAATT ACCTGAATCT	5040
TTTCATCTGT TAATTTTTCT GAACCTGCTA GCATAAGCTC TAGTGCCAAA TTCAAATTAA	5100
CGTCAGAGAA AACGATTACA TTTTCAAGTA CAGTTAGCCC CATCATAGCT GTAACCACAC	5160
TTGCACCAAA TAAATCACCC ATGAGGATAA ATTCATCTTC TGGCGCTTGT TCGGCTTTCT	5220
TTTTTAACTC TCCAAAAAA TCAACTGCAC TTTACCTGG ATACAGACTA TATGTATGAA	5280
CGGCAGCTGC AGAGTGTCCT GCAATCATTT CTAACGAGTG TTTAATCCCC TTAGAAAAAT	5340
CACCATGTGA AGCAATGTAA AACCTTCTCA TTTCATAAAC CTCCTTTATT TATTATATGA	5400
ACGAATGGCT ACTTTGTAGG ATTCGCTAAG GATATATTGC TCATTAAAAT ATAAAATTG	5460
ATTATCATT TCCACAACG TTTGATTAAT TTTGACCAAC GCTTCTTCTT CACTAATCGC	5520
TAATGCTTCT GCCAACTTTT CATCGGCCAA GGCAATTTC AATGATATTAT CAAAATAGCT	5580
AGCAATTACT CCTCTATATT GCCGTAAAAT ATCTAATAAT GAAGTATCTG TTAATTCTAA	5640
GTCCAACATA AACTGaTCTT CTTTTCTAAA ATAATCAATC TCGAAAATTG CAGGCATGTT	5700
ATCTAACGAA CGGAGGCGCT CAATAACAAT AACTTCTGTC GCTTTTTCAA CACCCAATGC	5760
TTCTGCAACA TGCCCTGTTG CTTTTTTCAA ACAAATTGAT ATGATCTTAG TTGTTGGTTT	5820
ACGATTCATT TTTAAACCTG ACTCAGTAAA ACTCCCACCA GCTGTCATGC TCTCAACATA	5880
GGATGGCATT GAAACAAATG TCCCTTTTCC ATGCCGTTTA ATTAAAATAT TTTCATCAAC	5940
TAATTGTTTA ATCGCATTAC GAACGGTTAC TCGGCTCACT CCGTAAATGG TCATCAATTG	6000
TTCTTCTGAA GGGATTTTTT CTCCTACTTG ATATTCACCA CTTCTAACTT TATTTAAAAT	6060
AGCATCGTAT AATTGTTGAT AAAgCGGTGC TTTTTTATTT GATTCTAATT TATCCATTAA	6120
TCATACACCT CATTTTAATA CGTCTTAATA CGTCTTAAAA ATAGCAAATA AAAAAGAGAA	6180
TGTCAAGAAA GAGTTTCCTT TCTTGACATT CTCTTTTTGC TTATTTTTCT TCAGAGTGAA	6240
CAATTAAGAC ATCGCAAGGG GCTTGACGAA TAATGTAGCT ACTGACGCTT CCCATGACCA	6300
CACGTTCGAn TGCGTTTAAT CCTGATTGTC CCACCATAAT TAAATCAACA TTGTATTTTT	6360
CAGGTAATTC ATGACTCATG ACTTCTTTGG GTGAACCAAA CATGGTAACT GTTTCAACAT	6420
TATGAAAATC AACGCTTTCA GCATACTTTT TACAATCGGC CATTAATTCT TCCGCAGCGG	6480
CTGTTTCTTG ATCAATTAAT TGATCATTTA ACGTCGAATA CCCCATCATT GTATACACTT	6540
GATTTTCTAA AATATTTGCA ACAATGACTC GGCTTCCGTT TCGACGAGCA ACCTCAATTG	6600
CACGTTCATA AGCTAAATTG GCTTGATCAC TACCGTCTAC GCCGACAAGA ATATTTTTAT	6660
ACATTTGTTC TTCCATTTTC GTCACCTTCC CGATGGAGGC TGATCTAAAA GGCTTCTTTG	6720

ACTAAGAAAC TAGAGTCAAA AATGAATTTA GGACCCTTGC GCAGAACGCC CTTTTGTCCA	6780
GCTGATTCTC TCTTAATTTA ATGGTTCCTC TGGTTCATGA AAATTAGCCT AAGCTAGTTA	6840
AAAATGCACT GATTTCTTCT TTTGTTTTTC GGTCTTTATT TACTAAGCGT CCTAACTCTT	6900
GACCGTTTTC AATTACGACA AAACTCGGAA TACCAAAAAT ATTCCACTCG GCAGCCACGT	6960
CAATAAACTG ATCACGGTCG ACTTCGATAA ATTGAAACGC TGGAAATGCT TCTTCAATTT	7020
CAGGCATGAC TGGTTTAATG AAACGACAGT CTCCGCACCA ATCAGCAGTA AAGAAGAAAA	7080
CATTTTACC TTTTCAACA TAAGTTGCTA GTTCTTCTAA TGAATTGGG <sub>a</sub> TAAATCATTT	7140
AAACATCCTC TACTTTCCGT AAACAATTGT ATTAAGTGT GATTTATCCA AGCCACGTAA	7200
TGCTTGTAAC AACATTTCCC GCGCTGCTTC ATAATCCGCA ATACTAAACA TTGTTTGATG	7260
CGTATGGATG TAACGACCAC AAACGCCGAT GACTGTACTT GGTACACCGT TATTAGTTGT	7320
GTGCGCTGCC CCAGCATCCG TTCCGCCTTT AGAAACAAAA TATTGATACG GAATATTGTG	7380
TGTCGCTGCT GTATCTAATA AATATTCACG TAAGCGAGGC AATGTGATCA AACCTGGGTC	7440
GAAAATCCGT AATAACGTTC CTTACCTAA ATGACCATAC GTCCCTTTTT TCGTGTGAAT	7500
ATCATCTGCC GCTGAACAAT CCACCGCAAA GAAGATATCT GGATTAAATT TATGAACAGA	7560
TGGTTTAGAA CCACGTAAAC CAACTTCTTC TTGTACGTTT GCTCCAGCAA TTAATGTGTG	7620
GCCTAATTCT TCATTTTGTA ACGTTTCCAA TGCATCTAAT ACCAACGTAC AACCATAGCG	7680
GTTATCCCAT GATTTACTGA TAATATTTTT GCCATTGGCT GTTTTAATCG TTTCTGTTTG	7740
TGGAACAATG GAATCCCCTG GACGAACGCC AAAGCTTTCA GCTTCTTCTT TTGAAGAGAA	7800
TCCTGCATCA AATAACACAT CAGAGACTTC TAATTGTTTT TGACCACTCG TACCACGCAA	7860
CAAGTGTGGC GGAATAGATG AAGAAATACA TGGATAGTTC CCTTTACTTG TTTTAACTG	7920
AAAACGTTGT GCTGAAACAA CATAGGGATT CCAGCCACCT AAAGGAACCA CTTGGAATAA	7980
ACCATTATCG TTAATTTGAG TTAACATGAA GCCAACTTCG TCCATGTGAG CTGCTACCAT	8040
TACGCGTGGC GCTGCTTCCA CTTTTGAACG TTTAATCCCA AAAATACCAC CTAAACCATC	8100
ATATTGGACA TCGTCCACTA ATGGGGTGAT ATGTTCTTTC ATGTACGCGC GAATGTCGTC	8160
TTCAAATCCG CTGGTTCCTT GTAATTCAGT TAATTCCTTG ATTCGTTGAA ATGTTTTTTC	8220
TTCCATAGAT GTATGTCCTA TTGCTTTTGA TTTGTTGGAA ATTACGCCAA CCAAGCGTCA	8280
CTTTTCTCT AAAAAGTCAC TTTATCCTGT CGATAAAAAT AGGTACTCCT TTCATATGAT	8340
TTCACTCTTA TTATAACGCA AATACTAGAA AATGCACATC GCTTCTCTT TTTAAACGTG	8400
ACAATGAGAA TGCCCTAAAA GAATGATTGA AAAACTTTTT TTATTTCTCA GCGGCCCTT	8460
CTCTATGGTA AAATAAGAGT GATACTGACA AAAGTAATAG TAAGCAGAAA CCGACCAAGG	8520
GAGGTCATAC AAATGAACGA AGAAAATGAG TTGACATACT TTAAAGGCGG CTTAGCGCTA	8580
GGCGTTGGTC TTGGTTTAGT AGGCGGCATT GCCTCAACCT TGTCTACCA TAAGAAAAA	8640
ACAATCTCCG CTGATCTCGT CTTAGAAAAT GTAAAAGCTG CTTTTTTAA AGAAGGACCG	8700

ATTGAAGGTT CGTGGATTGA GTTTGAGAAA AAGCCCTTGC GGAAATTCGC CATCCACTCC	8760
AAAACGTATA CTGGAGGAAT TTGCCGCATT GAAGACGACG GCATTGTTCA ATACGAATTC	8820
ACTGCAGATG CCTACACAGG AACTATCATT GACATTCAAC GATTAAAAGA TTA AAAAGAT	8880
TGTTACAAAA GCACTTATTT GCTTTTGTAA CAATCTTTT TTCTTCTCAT TTGTTTTTTG	8940
TAAAAGGGAT GTTTATTTTA TGTCAAAAA ATCAAATCAA TGATGAATAA ATACTCAAAA	9000
ATCGAATGCT TTCTGCATGG AAATCAGCTT AAATCTTTAC TTATCAACGT TTGTCCACTT	9060
TTTTCATAAA AAAGATTAGA CATTTTTTAA AAATTGATAA AAGAATAATC ATTTCTAACG	9120
ATTTTTTAAT TTCTCCTTTT CATTTTTTACA TTATACTGAG AGTATTGCTA AAAATGAAAA	9180
GGAGCGTGTC AAATGAAGAA ATCTGTTTTA TTTACTTCAT TACTTGTATT ATCAAGCTTA	9240
GCTTTAGCnG CcTGCGGCGG TGGCAGTGAC GATAAAGGAG CTAGCAACGG CGGCAGCGAC	9300
AACCAAGTAT ACACAATGGT TGAATCCCAA GAAATGCCTA GTGCCGATCC GTCCCTTGCG	9360
ACAGATGAAG TGAGTTTTAC CACTTTAAAT AATGTCTACG AAGGAATCTA TCGTTTAGAT	9420
AAAGACAACA AACCCGCGCC TGCTGGTGCA GCCGAAAAG CGACTGTTTC AGAAGACGGT	9480
TTAGTTTACA AAGTTAAATT ACGTGAAGAA TCAAAATGGT CTGATGGCAA ACCAGTTACT	9540
GCTGCAGATT ACGTTTACGG TTGGCAACGA ACAGTGGATC CTGCCACTGC TTCAGAATAT	9600
GCCTACATGT TTGAACCAGT AAAAAATGCT GAAAAATTT CTAAAGGGGA ACTACCTAAA	9660
GAAGAATTGG GCATTAAAGC AaTCAATGAT CATGAATTAG AAATCACTCT AGAAACAGCA	9720
ACACCATATT TTGACGATTT ATTGGCTTTC CCTTCTTTCT TGCCGCAACG TCAAGATATC	9780
GTTGAACGTT TTGGTAAaGA TTATACaAAG AGTAGCGATA AAGCAGTCTA CAaTGGTCCC	9840
TTTACGCTAA CTGAGTTTGA TGGTCCCGGA ACAGATACTA AATGGTCTCT AACTAAAAAT	9900
GAAGAATATT GGGACAAAGA GACGGTCAAG TTAGATAAAG TCGCTATCAA CGTGGTGAAA	9960
GAAGCCCCAA CTGCCTTAAA TCTTTATGAA ACTGGTGAAG TGGATGATAC GTATTTATCT	10020
GGCGAACTGG CTCAACAAAT GCAAACTCG CCTGACTTGG TCCAATTAAA AGCCGCTTCT	10080
TCTTTCTATT TAGAAATGAA TCAAGCAGAT GAAAAATCAC CATTGACTAA TGCAAACTTA	10140
CGTCGTGCTA TGTCTTATGC CATCGATCGC GATTCATTAG CTAAAAATAT TTTAGCTAAT	10200
GGCTCTCTTC CTTCAACAAGG CTTTCGTTCT GTGGATGTCG CGAAATCACC AAAAACGGGT	10260
GAAGACTTTG TTAAAGAAGC CGGCAGCGAC AAATTAGTCA AATACGACAA GAAAAAGCT	10320
GTGGAATACT GGAACAAAGC GAAACAAGAA CTTGGTGTTC CCAACTTAAC GGTTGATTTA	10380
ATGGTAGACG ACTCTGAAGG CGCTAAAAAA ATGGGCGAAT ATCTTCAAGG ATCACTATCT	10440
GATACTTTGG AAGGCTTAAA AGTAACTGTG ACGCCTGTCC CTATGGCTGT TCGCTTAGAT	10500
CGTACCTTAA AAGGGGATTT CCAAATCGCT GTTCGTGGTT GGAGTGCCGA CTATTCAGAC	10560
CCAATTAACT TCTTAGATTT ATTAGAAAGC TCAACTTCTA ATAACCGTGG ACGTTACAGC	10620
AATCCTGAAT ACGATAAATT CATTGCTGCG TCCAAAACCA CAGATGTAA CGATCCTGAA	10680

AAACGCTGGG AAGATCTAAT CAACGCTGAA AAAACAGTAA TTGCTGATAT GGGTGTGTG	10740
CCAATTTACC AAAAAGCAGA ATCACACTTA CGTGCACCAA ACGTGAAAGA AATTATTTAT	10800
CATCCAACAG GTGCTAAATA CGACTTCAAG TGGGCGTATA AAGAATAATC ACTCGATTAA	10860
ATTTAGAAAA AGATTGGGAA CAATAGCGAT TTTGGCTATT GTTCCAAaTC TTTTTTATAT	10920
AATCCCGATT AAAGAACCTA ACAACCCAAA AATCACAGTA AAAATAATTA AAGCAATCGG	10980
CGTGCCTCCT TTTTTCACAA TAATAAAACA AAACAGCGTA AACACTAAAG GTAATAAATT	11040
GGGCATGATT GTATCAaACA ATTGTGTTTG AAGGTCTAAC TTAGTGTCTT TCAAGCGAAA	11100
AACAAAAGTT GTCGTTAATT GTACACTCTT CGGAATCAAC GCGCCACCA CGGAAATTCC	11160
TACAATTGTT GCTGAACGAG ATAATTTTTT AGTACTTTCG CTTAAAACTG AAATTGCTTT	11220
GACACCCATT TTATAACCTA GATTAATTAA CCCAATTTTG GCAGCCATAT ACGGAATATT	11280
AAATAACAAC AAGAAAACAA TTGGTCCTAA GATATTGCCT TCTAATGATA GTTGTGAACC	11340
AATCGCTGCT GCAATCGGCA AAACGGTTAA ATAAAATAGC GCATCCCCAA TTCCGCCTAA	11400
CGGGCCCATC AACGCTGTTT TAATCGCTTG AATGGATTTT GGTGATTCTT TTGACTCTTC	11460
CATTGCAATA ATCACACCAT GCATGAACGT CACTAATTGT GGATTCGTGT TCATAAATTG	11520
TAAATGATTT TCCAAAGAAA GAGCTAAGTC TTCTTTATCA CGATGAATCT TTTTCAATCC	11580
TGGAATTAAT GAGTAGGTCC AGCCGCCTGC TTGCATTCTT TCATAATTAA ATGAAGATTG	11640
TAGCAGATAC GAACGGAACA TTATTTTTTT TAACTCTTTT TTTGTTATTA CTTTTTCGGA	11700
TATCATATCA TTATATGTTT CTTCAGATGC CATCTTCTTC AAACCCTCCA TTATTATTTG	11760
ATGCAGTGAC TGTATTATTT CTTTGTGCGT AAAATTCATA TAAAGCAATT GCGACAGCTA	11820
AACCTGCAAC CGCTAATGTA TCTAACCTTA AATAGGCAGC CATAACAAAC CCAAAAATAA	11880
AGAACGGCAC GTAGGATGTT TCCCACATCG TTTTAAATAA CATGGCGAAA CCTACTGCTG	11940
GCATCATCCC ACCTGCAACG CTCAAGCCAT GAATACTCCA TTCGGGTAGA ATAGAAATGA	12000
AACTAGAGGC TTTATCAGCA CCAAAGAAAA TACATAAAAA TACTACAAAA AAATAACAAA	12060
TGAATAAAAC GATTGGTAAA AACTAGACA TTCTTTCGAT GCCCCTTGTA TCTGCTTTTT	12120
TTGCATAAAC ATCAAATTTA TGCATAATTG GTGACATCGC TGTATATAAC AAGGTAATTA	12180
ACGCTTGCGC TGCAACTGCG AAAGGAACAG CCAAACcTAC GGTCACTTGT GGGTCAGATT	12240
TTGTGATAAT TGCTAACGCT GTTCCAATAA CTCCCCAAT TACTACGTTA GGTGGTTGTG	12300
CACCGCCGAT TGCTACTGCA CCAATCCAAA CTAACCTAG CGTTGCTCCA GTGATTAACC	12360
CTGTGTTGAC ATCTCCCAAT ATTAAACCAA CAATCAGCCC TGTAACAATC GGACGATGAA	12420
TATGTGTCTG TAAGACAAGC TTATCAACCC CTGCTATCCC AGCCCAAATT CCTACCAAAA	12480
TTGCTGTAC TAACATTTTT TTCCTCCTaT TAaCTTCGTC TACTTTTTAT AACTGTTTC	12540
ACCATTTACT AAGGTTTCTA GCACTTCATA TTTCTCATCA ATAATTGTAA AATTACCTAA	12600
AAGCCCTAGT TTGATTTCTC CCATAGATTC GTTCACATTC AACAAACGAG CTGGATTTTT	12660



CGTGACTGAA TTAATAGCAA TCACTTCTGG AAGTAATGCT TTTTCAACAA GGTGCGAAC	12720
CATATTATTT AAATGGTTCG TACTTCCAGC TAACTTGCCA TTGGCTAAAC GGACCACATT	12780
TTGTTTCATCA ATAACCATTT CGATCCCTTT TTCTGGTTTT GGATAAACGC CTGGTTGACA	12840
CCCTTTGGCC CAGATAGAAT CTGTCACAGC AATTAAATAG TCCTTGCCTT TTAGTTTGGC	12900
TAACACACGT ACTAATGAAT AATCAACATG TACCCCATCA GCAATAATTT CAGCAAACGT	12960
TTCTTCACTA TCCAAAGCAG CAACTACTGC GGTGTTTTT CGGTGACTGA TATCTTCCAT	13020
TCCATTAAAC GTATGGGTAA AACTTTTTGC ACCTGCTTCG ACAGCTGCCA TTGCTTGTTT	13080
ATAGGTTGCA GCCGTATGGC CAATTGAAAT GACTACATCA TGTTCATGAC AATATGTAGT	13140
CAAGGCATTC TCTACATCAT TTTCTGGTGC CAACGTCATT AATTTTATGT GCCCTTTTGC	13200
TAACTCTTGC CATTTTTTAA ATGTCTCAAC GGATGGTTTT ACTAATAACT CTGGATTATG	13260
AGAGCCACGA AATTCTTCAG AAATCATTGG TCCTTCCGCA TGAATCCCTA AGATTGTGC	13320
TCCGTTGGTT CCTTGGTCTT CCTCAATAAA TGAACCAATC ACTTCAAATG AATGTTCCAA	13380
GTCTTTTGA AATGTTGTTG ACGTTGTTGG CAAAAAGCT GTAATCCCTT CTTCTGGAAG	13440
GTAAGCTTGC CACTCTTTGA TAAATTCATG ATTGGCATGA TTGGCATCCC CACCATGCCA	13500
ACCATGATCG TGAATATCAA TAAATCCTGG AAGAATGCGC TTTTCTCCGT ATTCTTTGTC	13560
GACAGCTTTT TCATTGTACG GTAGAATTGC AACAAATCGT TGTTCGACA CTTCAACTTG	13620
AGCTGGTTGA AATTGTTTCGT TAATCCATAC GTTTTTACTT TGAATAATCA TTCTTCCAC	13680
TTCTTCTCTT CAAAATTATT GTCCAACATT ACTTTTTTAT GAAAATAATG CATCTAAACT	13740
TTCTTCTGGT GTTGACGGTA CACGTTGAAT CGAAAGTTTA ACACCTAATT CTTGTAGCTT	13800
TTTAAAGGCG TCTACATCTG TTTTATTAAC GGCAACAGCG GTAGCTACTT GTTTTTGcCT	13860
TCTTCCATAT GCATATTACC AATGTTTAAT TTTTCAATTG GCACACCACC TTCTACCAAG	13920
CGCAATGCAT CAACAGGATT ATCAACTAAC AGAAAAATCT TTTGTCTTTC TGAGGCTTTA	13980
TGAATGACCT CAATCGTTTT CTTAATGGAA AAAAAACGGG TTTGTACTTC TGCAGGGGCC	14040
GCCATGTTTA ACAATTGTTG GCGCAGTGTA TCCTCAGAAA TTGCATCATT TGCGACTAAA	14100
ATTAAGTTGG CGCCTAAAGA ACCGTTCCAT TGTGTTGCGA CTTGACCATG AATCAAGCGA	14160
TTATCAATTC TTGTTAATAA AATATTCGGC ATAGTGCTTC CTCCTCTTTT CTAATTGGTT	14220
AGATTCCGTC TGCTTCTTCT TGAGTATTTG CTTGTGGTTG TTCTAGTTTT TTATGAACAA	14280
TCCCATCTCT GGCTGAATCT GTGACCATTG AAACAAAATC TTCTACGTTT GTGCTGATTG	14340
TGCGTAATGC AAGCCCTTCA AGAACCATCG GTAAATTTGT TCCTGCTAAC ACTTCGACAT	14400
TTGGATAATT ACTTGCGACA GTCATTGAAA TTCTAAACGG TGTGCCACCT AAAAGGTCAG	14460
TTAAAATGAC CACTGCTGGA CACGTCTCCA GCATAGAGTC AACGGCTTTA CGAATGTTTG	14520
ACTCTAATAC ATCAAGCCCT GTATCATTGA AAAAAGGGAC TGCCTGAACA TTTGCTTGCT	14580
CTCCTGCAAT CATACTAACT GATTGATGCA TTCCAGTTGA AAATACCCCA TGCCCCGTTA	14640

AAATAAGACC TATCATTTAT TCTTCTCCC CTAACATTT AsTrmGCTTA CTATATATTG	14700
GATAAGCAGA ACTTATTTCT CAAAGAGTGT AACAATTCTT GTGACAATTG AGTATGCTTT	14760
CCCATGATTT TGGACGATGa TTTCTATTTA AGAACATGTC CCTATACAGA TACAAGGCGT	14820
CCCAACTTAA AGATGCTTTT GAAAAAATGC TCGACTACAA CCGTGAGTCC AAACGAGCCA	14880
ACAAATGTAC CCATATACAA CACAAGAATA ACCAAGCTGA CTGGTTGATt GGTTAAAAAC	14940
GGGTGGACCA GTTTCAGAC CACTTGAAAA ATAAAAACAT TCGCTAAATA ACTGCGATAC	15000
GCATAAgTGG CTATCCATTT AACCATTGGT AAAAGACGAA TCTTTTGGC TATTCCTTGc	15060
GTTCCCAAAT AAAAAACCA CCCGATAATC AATAAACTAT AGGCACTTTG AAGAGGATCC	15120
AAATAAGGTA GATCGTTCGG AGTCATCTGT GTTCCATCAA ATAAATTCCC TTGGACTGCC	15180
TGCCAAA AAA TAAAaAGaAA CAACAGAAAT AGTGCCACTC TCGTCTTTTT CAaGAAAGTa	15240
AAAATTTTCC CTTGATACTT CGCTGCGATG ACCCCGAAAA TACCATAAAT GAGATAACTA	15300
AGCCCTGAAC GATCTATTAA GTACCAAAGA GAAGTTTCTA CATATGGAAA AATCATTTTC	15360
TGGTACCAGA AGCACCAAAG GATATGCGCA CCGATTGTAA AAAAGACAAT TGGTAAAAATC	15420
ACTTTTTTGT TTTTGTAAAT TCTGGTTGCC AACCAAATAA AATACGGCAT AAATAATTGG	15480
AATTGCAACA TCATTACGGC ATACCATAGA TGTGCCGCGC CATCACCCT GAGAACTTTT	15540
CCTAGAAATC CTAGGGACGT TTCATAGGGC ATTGCTGTGT GTAAACTAGG AAAAGCGTAT	15600
AGATAGAGAC TGGACCAGAC CACGTAAGGA AGCACCAATT CATAAAATTT TTCCTTTAAA	15660
AATTCAAGAT ACGATTCAGC TTGACTCTTT AACACCATGT TGTAATAAAT TGCAAAAATA	15720
AACATGGGTG CCGAATATTT AGACATT	15747

## (2) INFORMATION FOR SEQ ID NO: 82:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1536 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TGCGTGAAAA GTATAAAGCT TTTGAACTGC CAnATAAGGT AGGAACTTCC CAGCAAAATA	60
ACCTnGTAAT TGACCATAAG CAAACAATGG CGAAACATTT CTTGAAAAAA ATTGTAACCT	120
GTAAAAATTCT TCATAAGGAT CCCCAAAATC CCATCGATTA AAATCAATCA CACCGACCTG	180
CCTAGTTGGC AAGTAAATAA AATTCCCAC ATGGAAATCT CCATGCTGAT AAACCACAGG	240
CCGTTGTTGC AATAACGGTA ATTGTTTTTC TAAATATGAA ATTATTGGCT GCTGGTACGG	300
CATACAAAAT TGAGATTCCT TATACTTATT TAGTTGGCTC TGCTTTTTTT CGTATAAATT	360
TTGCGCAATC TCTCTTGAC TAATCCTTTC AGGCAACAAA AGCTTATGGA TTTTGTTTAA	420
ATAGCGGCCT GCCTCTACGC CTAAGTTCAG TTGTCTTTTG GGAGAAAGAG CGGGTAAAC	480

ATCGCTTAAG TCTTCACCTT CTAGATACGT TAATAACATG TATCCTTGTT CACTTTCAGG	540
AATAGGTGCA CATAAAAAAG GTTTAGAACT GGGAAATCCT AACGCCGCTA CTTTTTTGAT	600
AAATTGATAT TCTTGTTGTT TGGCTAATAA CTCTTCTTGT TGAAAAATTC GCAATAAAAA	660
GGTTTCTTGA TTTTGGTAA CTTTAACTT TAAATCAGTT GACCAACCAC GGTGAATAGC	720
TTCAACTTTT TGCCATTGCT GAAACGCTGG AATCTTCTCA AACGCCCTCA TTACTTTCGC	780
CTCCATTTCC CTATTGGTTT TAGTATACCA AAGAAAAAAG GCTGGAGACA ATTGGGGATA	840
AATTGTCTCT AGCCTTCACT GGTTATCATT TTCTAGTTTT CCTTATATAA AtCGTTCCAC	900
TCGTTACTAA GAAAGCTCCT CCTAAAACAA CCCAGAAAGT ACTAGGTACT TCTGCACCTG	960
CTTTTGGTAA TACTTTTTTG GGAAGCTGTT CTTTTTCCC AATCATATTT TTATCAGAAA	1020
AAGCATGAGT ACTTTCGTGT GATTGGGTC CCTGATTGTC TGATGAATTC AGCGAATTAT	1080
CTAAGGATGA CTCACTAGTA GTTTTTCCAG ATTTGTCTGG GTTTTGGTTT TGGTTTGGT	1140
TTTGATCTTT TCCTTCGATT GGTCTTCAC TTGTTTTGT TTCTTTGTCT TTATCTGGTT	1200
CGCCACTTGT ATCTAGTTCT TTATCTTTAT CTGGTTGCC ACTTGCATCT GGGTCTTTAT	1260
CTTTGTCTGG TTCGCCACTT GCATCTGGGT CTTTATCTTT ATCCAGTTCG CCACCTGCAT	1320
TTGGTTCTTT GTCTTTGTCT GGTTCGCCAC TTGTATCTGG TTCTTTGTCT TtGTCTGGkT	1380
CTTCTACATG TGGTTGATTT ATCTTGTTGGT TATTTTCAA TAAAGCACCT AATTGTAGCA	1440
ATGTTGCGTC TTCATTTAAT GCACTATTAA ATTGAATCCC CAAAGGCAGT CCTTCTTTTG	1500
AAACGTAAGT TGGTACACTT AATGCTGGAT GTCCAA	1536

## (2) INFORMATION FOR SEQ ID NO: 83:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17250 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

GCCAAATTTG GTGGATCGTG TCCGCTTGGG AnnGCCTTAT GGGTGGACCC TACGTGCCAT	60
TTGGGTGAAA ACTAGGCCAT ATTTTAAAnG GAATATCCGT GCTTTCCTTG GCCACCGTTC	120
ATAATCCATT TCCACTGAAC CGCAAAAAGA ATCGTATCCA GAGAGATGGC AAGGCCAnAC	180
AGAGAAaTTT TyCTTkGGaA CAACTTCAAG AAACAATTCT TGAAAAAGCC AATCAACAAG	240
TAGGCAGTCG CCGGaCGACT AAAGAAGTCA TTAAATTTAC TACTAAAAAT AAAATTGGAG	300
ATAAAATGAC TCTTGGTACG AAGAGCTTGT TAATTTTGCT ACTTTGTATG ATTTTTTATG	360
TAAGTCTTAC CTTTGTAAG GGTATTTTAT TAACAACAGC TTTTATTAT CCAGCGGAAA	420
CAATTCAACG TATATTTGTT CCTGAGAAAG TAGCAGGGAA GGATACAAAT GCTGTGAAAa	480
CGACTAAAAC TAAACCAGCT TCAACAAAAG AATCTAGGCA GCAGGAGGAA GTTCAGCCTA	540
GCCCAACGAA CGTACCAGAA AATAATAATA GCGAACAAAGC TGTGTCAACG TATACTGTGG	600

AAGTTGGAGA	TTCAGTCAGT	TTGATTGCTG	AAAATCACGG	ACTCACTATA	GAACAACTTC	660
AAACTTTGAA	TCCAGAAATC	ATTGAGGTAC	CTATTTATCC	TGGACAAGTG	TTGAAATTGA	720
AAGAGGTGAC	TGAATAATGC	GCGAGGTAAA	ATTAATTTTA	GCTGaACAGT	TTTCaTACTT	780
AGGCsTTGCa	CAACGAATTG	CTAAATACAG	TAATAAAGCA	TTATATCAaa	GCCACATCTT	840
aGgAAATTTA	TGGCAAATTT	TAAGTCCTTT	AATTCAGTTA	GGTGTCTATT	ATTTTGCTTT	900
TGGGATTGCC	CTTGGTGGTG	CTCGAACCGT	AAAAGGAGGA	GTATCCTATA	TTGCTTGGTT	960
GATGGTAGGA	TTAAGTACAT	GGATTTTTTT	AAGTACAGTA	ACAAAACAAG	CTTCTAGTTC	1020
TGTTTATATG	CAAGTTGGCA	TGGTCTCTAG	AATGAAGTTT	CCAATAAGTA	TTTTGCCAAT	1080
GGTAAAATA	TTTTCTGAGT	TACCAAGTTA	TTTTGCATTC	ACTACGTTAG	CAGCGATTGT	1140
GTCAATCTCG	ACAGGTGAAA	AGATTTCTAT	TTATTGGATT	CAGCTACCTT	ACTATATTGT	1200
AGCGATGATT	ATCTTTTTAT	ATAGTTTTAG	TCTAATTAAT	TCAACTATTG	CTGCATTAAT	1260
TAGGGATTAT	CAAAATTTCT	TGAATTCTAT	TATCCAAGTA	TTAATGTATA	TGTCAGGGGT	1320
CTTTTGGGAT	TTGAGTACAA	AGAATTTACC	GTCTTGGTTA	TCAAAAATTC	TTATGTTGAA	1380
TCCTTACGCA	TATATTATTA	ATGGATTTTC	AGATACATTT	TTCTATAAaa	GATGGTTTTT	1440
TGAAGATAAG	CCCCAGCTTG	TTGTTTTTTG	GCTAACACA	TTAGTAATAT	TTATCGTAGG	1500
AGCTCATTTA	CATGTAAAGT	TCAGAAGTAA	ATTTGTTGAT	TATATGTAAC	ATGTGAAGAA	1560
GTAAAACCAC	CGTATACTGA	CAAGAATACG	GTGGTTTTAC	TCTTTAATTG	ACAATAGCCT	1620
TCTCTTTAAA	GAGAAGGCTA	TCTAACGAAT	CAATATTATT	ATTTTAAAGT	TGGATCAGAC	1680
TATTTCTAGT	ATCAAGTAT	TTTTCATTTA	AGCGATGTAT	TCCCTACCTC	TAATACATCC	1740
TTGTATCTCT	CTAGACACGC	CAgTTTTGAA	TAAAATCATC	AAAACAyTG	AATGGTTCGT	1800
TTAAAAATAT	TTTTCGTGTT	CTAATTATTA	TAACATAAAA	CTCGTCAACT	GACCACTAGC	1860
TGgACGAGTt	CATAATTTkt	CCAGATAAAT	AGACAGCCGT	TTCCGCCGA	CTGTCTAAGA	1920
TTAAGATAGA	TTTGTAAGA	AGTGgAATAA	ATGgCTTCCT	ACTATCTctT	TATAACakGA	1980
GGrAAAaGAAA	ACkGCaaaG	ATAGCTGTGT	TTTTTGAAAA	AAAGATCAGC	TGGTTTCCGC	2040
CAGCTGATCA	TCCGTTGAAA	AGAGATATAA	AGAAGTGAAT	AAAGCTTGCT	TTACCTCTAT	2100
TTTATACCAT	GTTTTGATTC	AAATAACTAT	TCAGTATTAG	TTATACTTAA	ATAATAAATA	2160
TGTATACGGA	TGTATTGTTT	TGAGGTATAG	TCATAGAGAA	AAGAGGGAAC	GTTTATGAAA	2220
AAATTTTGTT	TAAGTTATAG	CTCTGGTAAA	GATTGTTTAT	TGGCGATGGA	TCGATTGGTA	2280
CAAGCAGGAA	ArcAGCCAGT	TGCGTTAGTG	ACAACATTAA	GTGATGAAAT	TAATCGTTCT	2340
TGGTTTCACG	GCATACCAAT	CTCCGTCTTA	GAAGCAGCAG	CCGAAGCGTT	AGATTTACCA	2400
CTTGTTATAT	CACATAACAA	TGAAACAAAT	TACACAGAAA	AAGTGTTTGA	GGCACTGCAG	2460
GAAACCAAAA	AATTAGGTGC	AGAAACGGTT	TGTTTTGGCG	ATATAGATAT	TGAACAAAAT	2520
GGTGCTTGGG	ATCGCCAAGT	TGCGTTAAGT	GCTGGACTTG	AACCACAATT	ACCTCTTTGG	2580

CAGGAA <sup>---</sup> AATC	GCGAa <sup>---</sup> ACTT	GTGAAAGAAT	TCTTGGCAAA	AGGGTACACA	GCAATTATCA	2640
AAACGGTCAG	CAAAGAGGCA	GGCATTCCAA	TAAATTTTT	AGGAGAACCG	TTAAATGAAA	2700
CATTTATCAC	GTACCTTAAA	GAGCATCAAT	TAGATATTTG	CGGAGAAAaC	GGGGAGTACC	2760
ATACGTTGGT	AATTGACGGC	CcGCTATTTA	AAAAGAGGCy	CA <sup>---</sup> tTATTATm	CTcAGGAATT	2820
aTGAAgcCTT	ATGCTaTyCT	TgATtAt <sup>---</sup> tGA	TGCTTAAaTG	AAaGTAAaaG	GACACAAGCT	2880
GCCaCTTTTTt	CTAGGGATAT	gCTATAaTTG	mCTtAaGaaA	AGTAGAGAAA	GTGAGTGGTT	2940
AATTATGGAA	GAACTGTCT	TTTTTAATTT	AGGAAATGCC	CTTGCTTCAA	AAAGAGATCA	3000
AAAAGAATTA	ATTAAAGAAG	CCCAAATTGC	GAAAGATACT	CGAAAGATTC	CAGGAAA <sup>---</sup> ACT	3060
GATTATCGTT	GAAGATGAAG	AAAATGGTAC	CCATATTTTA	TTTGAATTAG	CTGATCAAAC	3120
AGAACCGTCT	GCTGAAACAA	AAGAATTCAA	AGTGAAAAAA	GTGATTGATT	AATAAAAAAG	3180
CAACGAATTT	TTCTAAATTC	GTTGCTTTTT	TATATTTTTT	TACACAAATG	ATACAAAAAG	3240
AATATCAAAA	TCAAATAAAT	GATTAGTATA	CTTGAAGaAA	CAAATAGGAA	GCAAAGCAAA	3300
CGTTTTGAGG	TGTTCTTTCA	TGGA <sup>---</sup> AAATCG	AGGAATTTTA	TTTGATAAAG	ATGGTACGTG	3360
TATTCGTTTT	GACACTCTTT	GGCAAGCAGG	TTTGCAAGCT	TGTTTTGAAA	CACTAAGTAT	3420
GTTaGCCCCT	CATCATT <sup>---</sup> CAG	CAGAAATAAA	AAAgATATTA	GCTATTCAGG	AGCAACGTTT	3480
TTTGCAAAAA	CATTTACTTG	ATGAAGTCCT	TTATCAGGAA	CTTTATCAGG	AATTGGCGCA	3540
ATTTGAGGAA	TTAGTCGAAC	AGGGAATCAG	CAGTCGATGG	CTGGAGCAAT	TTTTTTATGA	3600
TTATTTACGA	AAAAATCTGA	AAAAGATCGA	ACCGATTGGT	GATTTAA <sup>---</sup> AC	AGTTATTTCT	3660
TGAGCTGAAA	CGGAAGAACT	ATAAAATTGG	ATTAGCAACT	TCAGATACTT	TGCCAGCGAC	3720
TATGTTGATT	ATGGAATATC	TTGGTTTAAC	AGAAATGTTT	GATTTTATTG	CGACAGGAGA	3780
TCGTTACTTA	CCGAAACCAG	ATGCGGACAT	GCTCCAaGCC	TTTTGTCAGT	CATGTCAATT	3840
GAAGGCGACA	GAAGTAATTA	TGGTAGGTGA	TTCGCTCGTG	GATGTTTTTA	TGGGAACGTG	3900
TCATGGCAAA	GCAGGAATTG	GTGTTTTTAAC	CGGCAACTGC	CAGTCAACTG	ATTTTGAAAA	3960
GTTTGAAGTA	GCCTATTTCC	GTGATATCCA	TGAAATACCT	TATCAAGAAT	TATGGGAAAA	4020
TACCAAGAAG	AAAAAAATCT	ATTATTTTTG	ATATATGTAT	TGTGAAAAAC	GTTTCTTTTA	4080
TTCGAGTATC	TAAGCAAGCA	TACTAGTGTA	TTTCAATAAA	ATACCATGAT	GTTCATAGGT	4140
TTTATAATTT	CAGCTCACAG	ATTTTTGTTC	AATAAGAAAA	AAATTACAAC	TAAAATAGAA	4200
AACGGGCGTG	GGACAAAAAT	CACTTTGGAT	TTTTGTCCCA	CGCTCAAAAA	CTGATAAACG	4260
GCGGGAACAG	AAGCAAACT	TTCGGAAATA	AGCCGAAAGT	GACTTTCTCA	CACAAGTGTT	4320
CGGGATATCT	TAATCCTTAC	CTACGTGGTG	CTCGGAGTTA	AACACTTCTG	TCCTGACTTC	4380
GTTGTTAATA	TTTTACAATG	ACAAGTTTTT	CTTATTTTAA	TCCTGAACTT	GATTTAGAAA	4440
TTGCACTTCC	ATTAGAAAAA	GAAACAGTAA	CAGTAGCACC	TAAATCACCA	TCTAAGTTTT	4500
TCATCCAGAT	TAAAGTATCA	TTTTTTTCAC	CATTGATATT	TGTAGAAGTA	ATCCCTGTTG	4560

GGTCTCCTAA ATCTTTTCTA GCTTGTTCCTT .CAGAATATGA ACCATCAGTA GCAATGTTAT	4620
TTACTTGCTC TGCAGTAACT TTATCATGTT TTGCAACCTT CAAGCCACTT ACAGATTGTC	4680
TCGCTGCTTT GCCATCAGAG AATGAAACAG TAATAGATGC aAGAAGATCA CCACCCTTGA	4740
CATTTGTCCA AGATAGAGTA GTGGTAGAGA TGCCTTGGAT ATCTGTTGTA GAAGAAGAGG	4800
CTGGTTCACC TAAAAATTGCT TTTACTTCAT CTTCTGTAGA ACCGCCTTCG CTGCTGTTCa	4860
TAATATCACC GACATTAATT TTATCGTAAG TkGCTTTTAA ATCAGAGTTA TCTTTCGATT	4920
TATCTTCTGr TTTCTTTTCA GAGCTTtCTT TTTTGTATC ATTTTTTGAA CTATTCGTAG	4980
CTTTTGAAGA AGAGGCTTCT GTTTTCACTT CTTTGGTATC ATTGCTTGCT TTTTATCCT	5040
CTTTGCCAAG ACCACTACCA GCAATCCCCA GTACAATAAC TACTACAACC CAAAACCAAA	5100
TTTTCTTATA AAATGGTTTT TTTACTTTTC CTTCAATTGA ATCATTTTTT TTCATTTTCA	5160
CTTTCCCTCC CATAGTTTAG TTAATGATTT GTAGCGCTTA CTAAATAGTT GAAAACAATC	5220
CCCAATTAAA TAACCTTTGA CTACAAATCA CTTAGTTTAA ATGAAAATAA GTATACCGCC	5280
ATGTGTAAAA ATAACACACC ACAGTATACT TATTTTATCA TTATTCATA AAAATAACAC	5340
CATTATTTTT ATGAAATAAT TAAAAAGAT AAAAAATCTA AAAACTTCGA CCGCCGCCAC	5400
CGAATGTGCC GCCGCCAGTT GAGTGGGTGG TGCTACCACC ACCGCCCATT CCGCCACTGC	5460
CGCCATTGTT TTTAGGAATA CGACGCGTAG TGATGAAAGA GTTGGTTAAC TGATCTGTGC	5520
GGGAGGTTAA GTTTAAAGTT GTTTTTTCCC TAAAGGGATA TTGATAACTT GAAAATTTTA	5580
ATTGATATTT AGAAATATTA ATGCCTAAGA AGACCAACT GAGTATCAGC GCAGCAGCAA	5640
AAGCAATTAC CATTTCCAGC GGGGTAATGA CTTTATAACG AGTGATTTTA CCTGTTTCGC	5700
TGTCCACACG ATAGTGCCCC CCAGGAACCC CTTTATTAAC AAATGCTTGA GTTTCCTGAA	5760
CAAAGGTTTG AGCAGCCGCG AAATAATTTT CTTGACTCAT ATTATCCCAT ATTTTATCTA	5820
AGGTATCATC AATTCGTGCA TCTGTCATAT AATCAATCAT GTTTCAGAA GTAGAGATGT	5880
AGATTTTCCG TAAGTCCATA TCAATGAGAA AAAGAATCGC ATTTTGGTCC TGCCAACTT	5940
TATTTAAAAG ATAATGATCT GCATATTCTT GTTCATCGCC ATAGGTATTA TTATTTGTGG	6000
TTACAATAAA GACAGAGGCT TTTGTTTTTT CTTCTAAAGG TTGTATCTCT TGCTTTAGTT	6060
GGTTGATTTG ATCAGGTGTC AGTAATTGTG CGCCATCATC AATTGAATTT TCAGCCGCCG	6120
AAACGGGTAG GGCAAGGCCA AAGGTAAGAA GTATTAGGAA AAAAATAGGT AAAAGTCGTT	6180
TTTTCATAGG AAATAACCTC CTAATAAAAA GAGTGCTAAG AGTCCAGCGA AAATGCCTAA	6240
CGTTGCTAGC CCTAACGAG TGTAACATA TGGAAGTACA CCGCTTACTT TACCTGTTTG	6300
TCCATTCATG GCATAATAAT AAACTTTTTT TGATTGTTCC TGA CTGCGAT AGGTGACTAA	6360
CCAAATTGGC AAAAGCATAT AGTGATTATT TTCTTGGTTT AGAGAAATAT CCGTGCGAAC	6420
ATTTGTCAAC GTAGTATAAC CAGAAGCAGA GTCCCGTAAC AATGATTCCG AATAATCTTG	6480
TAATTCTTGT TGGACGTGTT TTTTGATTGC TTCATATTCA ATATCGCGTT TTTCTGCTTG	6540

GAAACCTGCT	AGGTATTGGC	TTTTGAAAGG	CACCGCTTGA	TCAATTAAAA	ATGGTTGAAC	6600
GCCTTCCACC	ATTTTTTGTT	GGACGTTTTT	TGACAAGGCA	TTTTTAACTA	ATTCTTTTAA	6660
CGATAATTTT	CCTGAACGTC	CCACATTGAA	ACGTTTTGTT	TCGGTGTATT	CAATATCACC	6720
TACGCGCCAA	ATGCGAATAC	TTGTGCCAGT	AGCTTGCAGT	TGTCCGTCGA	CTTCTGCATC	6780
AACGACCCAA	TAGGGGAAGT	AAACCCCACT	TAATTTATCG	ATTTGATTTT	TGCTGAAAAA	6840
ATCTCGGGGA	ATGAACCATT	TTTTCTTGGT	CCAAGCGAGA	AATTTTTCAA	TCGCTTCGTC	6900
TTTTTCAATT	GTAAAGGGTA	AGACATTTTC	TGGTAAAAAT	TCGCCACTTA	ATCGACCAGC	6960
CAAGACAAC	GGATTATGGC	AATAATAACA	GTATGTGGCA	GCCGTTGTTG	CATCGGTGAC	7020
AATTTGAGCG	CCACAGCTTG	GACAATTAAA	AATTTCCATT	GTGCCTACAT	CTGCAGTTGT	7080
TTCAATCTCA	GTGGTCTCTG	TTTGTTCGCT	ACGGAGGATT	CTTCTTCCGC	TGTGAAAGTC	7140
AAATTTTCCT	CGGCAGTCGT	TGTTGAGGGA	TTCTCTAAAT	GCGCCTCATG	TTGCTGTTGC	7200
TCATAACGAG	TCACTTCTTC	TTCCGTGTAA	ATATTTAAAC	AATATTACAC	ATGAAATTTT	7260
TGATCTTTGG	GATTAAATAA	GAGCGGGCCG	CCACAGTTGG	GGCATTATATG	GGTAAATGTT	7320
GTTTCCATTG	CAACCGTCCT	TTCTAATAGA	GTGAGTAAAT	AACAGGAATC	TGATTCAAAG	7380
AAAAAGGGTA	GACTTGTGCC	TGACCCTTTT	CTTTAATTGT	TCACTACTTG	CTTATAAAGG	7440
AACCCCTTGG	AATGGTTTAC	CACAATGTGG	ACAGAACTTA	GGAATACCAT	TACTTAAGTC	7500
AACAATTTTCG	CTACATTCAC	TACATTTTCT	TTCTAATTTT	GGTTTCTCGT	TTGTTGGTTT	7560
TGGTGTTACCA	CAGTTTGAAC	AAAACCTCCC	AGTGTTTTCA	GTCCACACATT	GTGGGCAAGT	7620
CCAAGTGCT	GTAGCCCTT	GTGCGTTTTG	CTGGTTTTGT	TTTCTTGCA	TTTGTGTTG	7680
GTTTGTGTTGA	GACGCTTGCG	CAAAGTAGTT	GCCATTGCGA	TTCATGCCCA	TACCAACGCC	7740
CATAAAGCCG	CTCATAGCGC	CACCTTCATT	TTTACCAGCA	GCCTCCATAC	CACGAGCCAA	7800
GGAACCTTGA	ACATAGCCTT	CGCGAACACC	TGGATCGCTG	AGCATAGCAC	CTTCATTTTCG	7860
CATGTTGATT	AATTTTGTTG	AATCATCTGT	ATAAGAAATG	CTCGCAACTG	CTACAGAAAC	7920
AATTTCCATG	CCGCGATTGG	TGCGCCAATC	CTCATCAAGC	GCAGTTTGCA	TATATTTGCT	7980
TAATTCAACA	CTTTTTGAGG	GAACATAAGA	AATACGTTGC	CCATCAGCAG	ACATTTGATT	8040
GATTGCTGCT	TGCAATGCTG	TAATGAACTC	TGCCAAATAT	TGTTCAATTA	TATCATTAAT	8100
ATCTACTTGC	GTTTGGTTTT	TAGGAATCGC	ATTAGTATAA	AATAGCAAAG	GATCTGTAAT	8160
TTTAATTGAG	TAATTTCCGT	GTGCACGTAA	GAAAAGTTTCG	GCATTATAAA	AATTATCAAA	8220
ATAGTTTAAT	GGCGCAGGTG	TTCCAAATTT	AATGCCTTTA	ATTTCTTGCA	AATTAATATA	8280
AAAAACTTGT	TGTTTTTGTTG	GAGTTACGCC	ACCAAATTTG	AAACGACTGA	AGGTCTCTGC	8340
AATCGCATCT	TTTAAAGAAC	CGTTAAACAT	TGAAGGCGCA	GAATCATTTT	TGACTGTATA	8400
ATAGCCTTCT	TCAGCAGTGT	AATCGATGAT	TTTGCCACCA	TCGACAAGAA	GCATCATCAT	8460
GTTGGGATAG	ACATGTACAA	CAGAACCATC	TGTTAAAACA	TCGTCCGTTT	CTTTACGATT	8520

AGAACCACGT TTATCATCTT TACGTACAAA AACACCTTTT GTCATGACGG TTGTGTCCCC	8580
CATATTATCA GGTTC AATAA CCTCTAGCCA TTGATCGGCC AAGCCACCGC CAACTGAACT	8640
TGTTGCTGCT TTAATAAGTC CCATGTTTCAT TCCTCCTAAT AAAAACTAA TTGTTAAAGC	8700
TAGTATACCA CACAATGAAT GCAGAGAAAT CCAACGAAAG CCTGATGAAA TGGAAAAAAT	8760
TGATCAAACCT TTATTAAGAA ATTAAAAAAA GTATAGGTGA AACACAAGCA TTTTAAAAAA	8820
AGTTACGTAT AATAGAAGAA AAGACATATT ATATAGTTAG GGGATTATTG AATGGAACGA	8880
AGCAATCGTA ATAAAAATC TTCAAAAAAA CCACTTATTC TTGGTGTTTC TGCCTTG GTT	8940
CTAATCGCTG CTGCCGGTGG CGGGTATTAT GCTTATAGTC AATGGCAAGC CAAACAAGAA	9000
TTAGCCGAAG CGAAGAAAAC AGCTACTACA TTTTAAACG TATTGTCAA ACAGGAATTT	9060
GATAAGTTAC CGTCCGTTGT TCAAGAAGCT AGCTTAAAGA AAAATGGCTA TGATACTAAA	9120
TCTGTTGTTG AAAAATACCA AGCAATTTAT TCAGGGATTC AAGCAGAAGG AGTCAAAGCT	9180
AGTGATGTTT AAGTCAAAAA GCGGAAAGAC AATCAATACA CATTACCTA TAAATTATCG	9240
ATGAGCACGC CTTTAGGCGA AATGAAAGAT TTGTCTTATC AATCAAGTAT CGCCAAAAAA	9300
GCGGATACCT ACCAAATCGC TTGGAAGCCA TCTTTAATTT TTCCAGATAT GTCAGGAAAT	9360
GATAAAATTT CGATTCAAGT AGATAATGCC AAACGTGGAG AAATTGTCGA TCGTAATGGT	9420
AGTGGGCTAG CAATTAACAA AGTGTTTGAC GAAGTGGGCG TAGTGCCTGG CAACTCGGT	9480
TCTGGCGCAG AAAAAACAGC CAATATCAAA GCTTTTAGTG ATAAATTCGG CGTTTCTGTT	9540
GATGAAATCA ATCAAAAGTT AAGCCAAGGA TGGGTCCAAG CAGACTCCTT TGTACCAATC	9600
ACAGTCGCTT CTGAACCAGT GACAGAATTA CCAACAGGGG CTGCGACAAA AGATACAGAG	9660
TCACGTTATT ATCCGCTGGG GGAAGCAaTG CGCAATTAAT CGGGTATACG GGCACCATTA	9720
CTGCAGAAGA TATTGAAAAA AATCCAGAGT TAAGTAGTAC GGGCGTAATT GGTA AAACTG	9780
GCTTAGAACG AGCGTTTGAT AAAGAATTAA GAGGGCAAGA TGGTGGCTCA CTAGTTATTT	9840
TAGATGACAA AGAGAAaTGTC AAAAAAGCTT TACAAACCAA AGAGAAAAAA GATGGTCAAA	9900
CGATTAAATT AACGATTGAC AGTGGCGTAC AACAACAAGC TTTTGCTATA TTTGACAAAC	9960
GCCCtGGTAG TGCGGTCATT ACCGATCCTC AAAAAGGCGA TTTATTAGCA ACGGTTAGTT	10020
CACCTTCGTA TGATCCTAAT AAAATGGCCA ATGGCATTTT TCAAAAAGAA TATGATGCTT	10080
ACAATAACAA CAAAGATTTA CCATTACAG CACGTTTTGC GACAGGCTAT GCACCAGGCT	10140
CTACCTTTAA AACAATTACC GGTGCCATTG GTTTAGATGC AGGGACATTG AAGCCAGATG	10200
AAGAGCTAGA AATTAATGGT TTGAAATGGC AAAAAGATAA ATCTTGGGGC GGCTATTTTG	10260
CGACACGTGT AAAAGAAGCA AGTCCAGTCA ATCTGCGTAC CGCTTTAGTT AATTCAGATA	10320
ATATTTATTT TGCCCAACAA ACATTACGGA TGGGAGAAGA CAAATTTAGA GCGGGTCTTA	10380
ATAAATTCAT TTTCGGTGAG GAACTAGATT TACCAATTGC CATGACGCCT GCTCAAATTT	10440
CAAATGAAGA TAAATTTAAT TCTGAAATTC TTTTAGCAGA TACGGGCTAC GGTCAAGGTC	10500



AATTGTTAAT	TTCCGCAATC	CAGCAAGCTA	CCATGTATAG	TGTTTTCCAA	AATAATGGAA	10560
CATTAGTCTA	TCCAAAACCTA	GTTTTAGACA	AAGAAACGAA	GAAAAAGAC	AATGTAATCA	10620
GCGCCAACGC	AGCGAATACA	ATTGCCACAG	ATCTTTTAGG	TAGTGTGGAA	GATCCTTCAG	10680
GTTATGTTTA	TAATATGTAC	AATCCGAACCT	TCTCATTGGC	TGCTAAAACA	GGAACAGCTG	10740
AAATAAAAGA	TAAACAAGAT	ACAGATGGCA	AAGAAAACAG	TTTCCTTCTA	ACGTTAGATC	10800
GTAGTAACAA	TAAATTCTTA	ACAATGATTA	TGGTTGAAAA	TTCAGGAGAA	AATGGTTCTG	10860
CAACAGATAT	CAGTAAACCA	TTAATTGATT	ACTTAGAAGC	AaCCATTaAa	TAAaAAGAGa	10920
AAATGAaCGA	AcGAGGTTtC	TGTAaaaaAGA	AcCTCGTTTT	CCCTTTTTTAG	AAAAAGGAAA	10980
GTGTGCTACA	ATGAAGCGAC	TAAATTGTAG	AAGGAAAGGG	ATTCGTGTGC	GAATTTTAGC	11040
AATAGATACA	TCTAACCAAA	CATTGAGTAT	TGCTGTGTGT	GAAAATCAAA	AAATTCTGGG	11100
AAGTTATACA	GCAACGGTTA	AAAGAAATCA	TAGCTTAACA	TTAATGCCGG	CAATTGACTA	11160
TTTGATGAGC	CAATTGAACT	TGGCGCCAAC	GGCAATTGAT	CGTTTTGTCTG	TAGCAGAAGG	11220
ACCTGGATCA	TACACTGGCT	TACGTTTAGG	CGTGACGACT	GCCAAAACAT	TGGCCTATAC	11280
TTTAAAGAAA	GAGCTAGTAG	GCATTTCCAG	TCTACAAACG	TTAGCTGCCA	ATTGTGTTGG	11340
ACAACTGGT	TTAATTGTGC	CTCTTTTGA	TGCGCGTCGT	AAAAACGTCT	ATGCTGGTGC	11400
TTATCGGTTT	GTCAATGGCG	TGTGGCAAAA	TGAGTTACCG	GACCAACATA	TTTcCCTGCG	11460
GGAGTTATTA	GAACAACTAA	AAAATGAACC	AAATCTTTTT	TTTGTTGGTG	AAGATGTTGA	11520
AAAATTCACT	GAAGAAATAG	CTCAAATCAT	TCCTCATGGA	GAAATTTGTG	ATGTTCCACA	11580
ATGGCAAATT	CCTAATGCCG	CAGTTTTGGC	TGCGTTAGGT	AGTGTGGCTG	AACCTGTTGA	11640
GAATGTTTAT	GGGTTCTTAC	CTGCGTATTT	GAAAAAGTA	GAAGCGGAAG	AAAATTGGTT	11700
GAAGACACAC	ACGCCTAATG	GAGAAAGCTA	TGTGGAAAAA	CTTTAATTTT	GTGACAAAAT	11760
ACTTTACAAA	AAGACGGCAA	CGGTATATAA	CTAAAACCAT	TGCAATGAAC	CAGCAAGAAT	11820
TTCAATTGCG	TGAAATCACC	TACCAAGATA	TTAAGGATTT	ACTAGCAATT	GAACGTGAAG	11880
TTTATGCTGG	GGAACCTCCT	TGGACGATGT	CTGCTTTCAT	GGCAGAACTT	GGCTCAAAAG	11940
CGCCGCATCT	CTATTTGTTA	GCGGCAATTA	ACGGAAAAAC	AGTTGGCTTT	ATTGGTTGTC	12000
GGATTCAAGG	AACGGATGCA	CACATTACGA	ATGTCGCTGT	TCACACAGCA	TACCAAGGCT	12060
TAGGTTTAGG	TCGTTCTTTA	ATAGAAGAAA	CGCGAATTTT	TGCCAAGAAA	AATCGGTGTG	12120
AAACCTTATC	GCTAGAAGTA	CGAATGAGTA	ATGTTAAAGC	CCAACGTTTA	TATCGAAAAA	12180
TTGGCTTTGT	CTCACAGACG	GTGAAAAAAG	GCTATTATGA	CGAAAAATAAT	GAAGACGCAC	12240
TAGAAATGAT	AATTGATTTA	AAAGAAGAGT	GAAGAAAAAA	TGTTAGTAAC	AAAAAAAGAA	12300
ATATCTGTAG	CAAATATTGG	TGAAGAATTA	TGGCGTTTTA	GTAATTCTTC	TTATACAACT	12360
GGCTCCCCCT	GGaCAGAAAC	ACAATTCGCG	GAAGATTTTG	CTCAGGaAAA	TAGCGAATAT	12420
CTTTTTTTAG	TTGAAAATGG	CCAGTGGTTA	GGCTATATTG	CCTATCATTT	TATTTTAGAT	12480

GAAGCAGAAA	TTAGTCACGT	TGTTGTGAAT	GGTCAGAAAC	AACACCAAGG	AATTGGTTGT	12540
ccAGTTAATG	aAAGCTTTTA	AAGAATATGT	TAAAAGTCgT	GACATTACAC	AAATTTTcTT	12600
AGAGGTCAGA	GAAyCGAATA	TACTAGCyCm	AAAACtTTAk	GrGAAAaCAG	GTTTtCGGAA	12660
AGTAGCTGTT	CGAnAAAATT	ATTATAAAGA	ACCGCAAGAA	AATGCCTTTG	TAATGTGTGC	12720
GAAATTAAGG	AAGGAAGCCC	AATGACAATT	TTTACAAAAG	AACGTAAATT	ATTATTAGCT	12780
GTGGAAAGTA	GCTGTGATGA	AACAAGTGTG	GCTGTCATTG	AAGATGGCGA	CAAAATTCTT	12840
TCCAATATCG	TAGCTTCGcA	AATTAAGAGC	CATCAACGTT	TTGGCGGAGT	CGTTCCGGAA	12900
GTTGCGAGTC	GTCATCATGT	TGAACAAGTG	ACGATCTGTA	TTGAAGAAGC	ATTAACAGAA	12960
GCAAAAGTCA	CACCAGAAGA	ACTCAGTGGC	GTTGCTGTGA	CCTATGGGCC	TGGGTTAGTC	13020
GGTGCTTTAT	TAATAGGACT	TTCTGCAGCA	AAAGCATTTG	CTTGGGCGCA	CCAATTACCA	13080
CTTATTCCAG	TCAATCATAT	GGCTGGGCAC	ATTTATGCTG	CCCGTTTTGT	GGCACCGCTC	13140
GAATTTCCAT	TGATGGCCTT	ACTCGTTAGT	GGAGGACATA	CGGAATTGGT	CTACATGAAA	13200
GAAGATGGCT	CTTTCGAAAT	TGTTGGTGAA	ACACGTGATG	ATGCAGCTGG	TGAAGCGTAT	13260
GATAAAGTCG	GACGTGTGTT	AGGGTTGCCT	TATCCTAGTG	GCAAAGAGAT	TGATGCTTTG	13320
GCCCATGAAG	GCACAGATAC	CTATCAATTT	CCACGGGCGA	TGCTAAAAGA	GGATAACTAT	13380
GACTTTAGCT	TTAGTGGTTT	GAAAAGTGCG	TTTATTAACA	CGGTTCATAA	TGCAGAGCAA	13440
CGTGGTGAAG	CACTGTCTAC	AAAGGATTTA	GCCGCaAGTT	TyCAGGCGAG	TGTGGTAGAA	13500
GTATTAGTGA	CTAAAACCAT	TCGTGCTTGT	CAAGAATATC	CTGTTAAGCA	ATTGCTAATC	13560
GCTGGGGGCG	TTGCTGCGAA	TCAAGGATTA	CGAGAAGCAA	TGAGACATGC	AATCAGTGaA	13620
CAATTACCAG	CAGTGACATT	ACTTATTCCG	CCATTAAAAC	TTTGTGGGGA	TAACGCGGCG	13680
ATGATTGGCG	CTGCCGCCCT	TATCGAAGCT	GcAAAAGAAT	CACTTTGCTT	CGTATAACTT	13740
AAATGCAGAA	CCTGGCGTTA	GTTTTATGAC	AATTAGTGAA	GAAGGCTAGC	CAGCTAGACA	13800
AAAATTTGCT	ATAATGAAGC	CAGTTTAACA	ATGACATTCG	ATGATAAAGA	GAGTACTTAT	13860
TTCAACGTCA	GCTTAGCGAT	CTTGGGAAGG	TGGGAGCCAA	GAGTGATCGG	AATGAGGAAG	13920
CGCACTTTTA	GAATTCTGCT	GAAAGCAGAC	GGCAATCACC	GTTATCGATG	AAGCAGGTCT	13980
ATCCAAGTAG	TATAGACAAT	TAGAGTGGTA	CCGCGGGTAA	ATTckGGGnT	CTGACAAAAC	14040
AAATTATTTT	GTTTTGTCAG	AGACGTTTTt	GTTTTTTTAA	TACTTACAAA	TTAnTCAGGA	14100
GGGTTCTTAT	GAaCAACAAA	GAAATCGTAG	CAAAAGCGTT	ACATGATGtT	TTAAATGAAG	14160
AATTAACGAT	GGATCAAATC	GAACAATTAT	TGGAAAaTCC	TAAATCTGTT	GATCATGGAG	14220
ATGTTGCTTT	TCCAGCCTTT	TCATTAGCTA	AAATTTATCG	TAAAGCACCA	CAACAAATTG	14280
CAGCAGAATT	AGCAGAAAAA	ATTGATGGCA	CAAAATTTGA	AAAAATTGAA	GTGGTGGGAC	14340
CTTATTTAAA	CTTCTTCATG	AATAAAGAAG	CAGTCAGTCA	AGCTGTCATT	GGCGAAGTAG	14400
TCAAAGAAAA	AAATAACTAC	GGAAACAGTA	CCTTTGGCAA	TAATGGCAAT	GTTCCATTG	14460

ATATGTCTTC	ACCAAACATT	GCTAAACCAA	TTTCTATGGG	GCACTTACGT	TCTACGGTAA	14520
TCGGGAATTC	AATTGCCTTT	ATTTTAGAAA	AAATTGGGTA	TCAGCCAATT	CGGATTAACC	14580
ACTTAGGTGA	CTGGGGTACA	CAGTTTGGTA	AGTTAATTGT	AGCTTATAAA	AAATGGGGTT	14640
CTGAAGAAGC	AGTCCGCCAA	CAACCAATTA	ATGAATTATT	GCGCTTATAT	GTGCAATTC	14700
ATGAAGAAGC	AGAAGAAAA	CCAGAATTAG	AAGACGAAGC	ACGTGCATGG	TTAAAAAAT	14760
TAGAAGAAGG	AGACCAAGAA	GCTAATGAAT	TATGGAAATG	GTTCCGTAGC	GAGTCTCTAA	14820
AAGAATTTGA	TAAAATCTAT	TCCATGTTAG	AAGTTGAATT	TGATTCTTAT	AATGGCGAAg	14880
cyTTTTATAA	TGATAAAATG	GATGAAATTG	TTACTTTATT	AGAAGAAAAA	CATTTATTAA	14940
CAGAAAATCA	AGGTGCAGAA	ATTGTTGATT	TGACAGAATA	TAATTTAAAT	CCAGCACTTA	15000
TTCGTAAATC	AGATGGTGCG	ACGTTATATA	TCACACGGGA	TTAgCTgcT	GCGCTTTACC	15060
GTAACCGGAC	GTATGATTTT	GCCAAATCGC	TTTATGTAGT	CGGCAACGAA	CAAAGCAATC	15120
ACTTCAAACA	ATTAAAAGCA	GTCTTGAAAG	AACTAGGGTT	TGACTGGTCA	GATGATATGG	15180
AACACATTCC	GTTTGGTTTA	ATTACTCAAG	GCGGTAAAAA	ATTATCCACA	CGTAAAGGGA	15240
AAATCGTTTT	ATTAGAAGAA	GTATTGAATG	AAGCTGTTAC	TTTAGCAGGA	AATCAAATTA	15300
ACGAGAAAAA	CCCAGACTTA	GCAATCGTG	AAGAAGTGGC	GAAACAAGTA	GGTGTAGGTG	15360
CAGTTATTTT	CCATGACCTT	AAAAATGATC	GTCTAAATAA	TTTCGATTTT	GTATTAGATG	15420
AAGTTGTTTC	TTTTGAAGGA	GAAACTGGGC	CATATGTTCA	ATATACACAT	GCTCGTGCGA	15480
TGAGTATTTT	ACGTAAAGCA	AACTTTACAC	CAGACGCTAC	ACAACGTTAT	GCACTAAATG	15540
ACAAGGATAG	CTGGGAAGTC	GTGAAGTTGT	TACAAAAATT	CCCTGAAACT	GTGATGCAAG	15600
CTGCTGAAAA	ATATGAACCA	TCTGTTATTG	CGAAACATAG	CATTCATTTA	GCACAAGCGT	15660
TCAATAAATA	TTATGCCCAT	GTTCCGATTT	TGGATGAAGA	TGCTCAAAAA	GAGGCACGCT	15720
TAGCGTTAGT	TTATGCAGTA	GCTACTGTTT	TAAAAGAAGA	TTTACGTTTA	CTAGGGTTAC	15780
ATGCACCAGA	AGAAATGTAG	AAGTTTTTGT	GAAAATTTTA	CTGCGTTGAA	CGCTAAACTT	15840
ATTAAGCAAA	TATTAAAGTT	AAGTTACAGA	TTGAGAAATT	CATAGTAATT	AAATAGCTGG	15900
GGTGCTATAC	TCGAGTCAAG	TTATAAAAGG	ATTAGAAAAG	GTGGCACCCA	ATTATATGAC	15960
AGAAAAATGG	CGTGAAGACG	aAGAATATCT	TTCTTATGTA	GAAGATTTAT	TAGcAACCGA	16020
AGAAGTCCAA	AAATTAGCAG	AATACACACA	ACATATGCAT	TCAACTCGTT	TAGAACATTC	16080
AATTAGTGTT	TCCTATCATA	GCTATAAATT	AGCAAAGCGT	TGGAACGGCA	ATGCTCGTGC	16140
AACAGCTCGT	GCTGGTTTAT	TGCATGATTT	GTTTTACTAT	GATTGGCGCA	CAACAAAGTT	16200
TGACGAAGGC	TCTCATGCCT	ATGTTCaTCC	aCGAATTGCA	GCCAAAAATG	CTGAAAAAAT	16260
TACAGAGTTG	TCTGATTTAG	AACGAGACAT	TATTATTAAA	CATATGTGGG	GCGCAACTAT	16320
TGCACCACCA	AAaTACAAAG	AAAGCTACAT	TGTAACCTTT	GTTGACAAAT	ATTGTGCTGT	16380
CAAAGAAGCG	gCGTTGCCGA	TGACGACAGC	CATGAAAACC	AAATGGCGTC	AGTATTTTGG	16440

TCAACAGTCA ATCTAAAAAA AGCTTATCCG TAAGGATAAG CTTTTTTTGT TTGACGTTGA	16500
CTAAAAAGTC TAGCATTTCAT CTATTTAAAA TAAATATTAA AAGTAATTTT GTTCTGTAA	16560
TTTCTTTTGA GCGTATCGAA AGACTGCTTC TTGATTTTCA GTTGTTAACT GTTTGTAAAT	16620
GTTCATAATA TTGGTTGGGA CAACAAATGG CGTTTCTTCC GTTTCAGGAA TACCTAAGAT	16680
TTCAAGAGGT GATACATCTA AAATTTTAGA TAAGGCGACA ATTTTATCCC TCCCCATATT	16740
TTCAATCATT CCAGTTTCCC ATTTCTGAC GGTACTTTT CCAACACCAA CAAAATTTCC	16800
AACTTGTTCT AATGTAAGAT TTTTCTCCAT CCTTTTTTGC TTCAAAAAAT TTTCTTTCAA	16860
AAAAACCACC TCCGTAAGAA CATTCTAGCA CAAAAGTGT CGCAAAAGAA ACATGTTTTT	16920
tTCAAAATGA AAGTAACGAA TATAATAGnA AAATCAGCAT TTTTAAATGA TATTGACGTT	16980
TTTTTTCACAA AATGATATAA TTTTAAGTGT CTTATTAGAC ACTGAAAmGG GGGAGAGTAA	17040
AATGGAATCA AATTTAATTC ACTCAAAATT GaAAATGAAA GGAAAAATG TTGaGTGGTT	17100
GGCAAAATGAA ATGAcTGCaT ACGGGCAACC AATATCCGCT TCAACGATTT ACAAAAAATT	17160
AAAAGGAGAA GTTGTTTTCA AAGCATCGGA TATTAAATTG ATTTACAAT TGTGGATTT	17220
GAATAATGAA GAAATTATGG ACTATnTTTT	17250

## (2) INFORMATION FOR SEQ ID NO: 84:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7947 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

AGATTAnCGG AAnCTACCTA TCTGAACGTA CAGATAATAA TATTACmATG CAAGCAATTC	60
GTGATCATTa TGAAGTATCC AATCGATTtG GCGCAAAAGA AGTCTTGGCT TTTTCTTCTG	120
AACGAAAATG GGGAGCTATT GAATTTCCAG AAATTGGGAC AGTTTATTTA GGGGCTCCTG	180
AGCGTTTGGT AGACGATAGT CGGTTGCCAG AAGCTGTTTT TACAGCACAA GAGAATGGGT	240
ATCGCGTCTT GATGTTAGCT ATTGCGGAGC AACaACCGTT AAACGAAACC AAAATGCCTT	300
ATTTAGAGCC ATTAGCAATT CTAGAAATTG ATGATCCAAT TCGCCAAAAT GCCAAAGAAA	360
CACtGGCTTA TCTAAAAGAA GAAGGAATCG ACTTAAAAGT AATTTCTGGT GATAATCCGG	420
TGACTGTTTC AAATATTGCT CGTCGTGCTG GGCTACCCGG TTATGAGTCT TATATTGATT	480
TGTCGACCAA AACGACAGAA GCAGAAGTGC GTGAAGCAGT TCAGCAATAC ACAGTGTTcG	540
GACGTGTATC GCCTCAGCAA AAACGAACCA TTGTGCGTGA GTTGAAAGAC ACAGAGCATG	600
TGGTGGCGAT GACCGGCGAT GGCGTTAACG ATGTTTTAGC GTTACGAGAA GCGGATTGTA	660
GTATCGCCAT GGCAGAGGGA GATGGCGCTA CTCGGCAAAT TTCTAATCTA GTTTTGCTAG	720
ATTCTGATTT TACAACGTTA CCAGATGTCT TGTTTGAAGG ACGACGTGTG GTAAATAATG	780

TAACCCGAGT	TTCCAGCGTT	TTCTTCATTA	AAACGATTTA	TTCGTTTATT	CTCTCAATCA	840
TTTGTGCCTT	AACGGCCATT	GCGTTCCCGT	TCATTCCGAT	TCAAGTGACC	TTGATTGATT	900
TAGCAATTGA	AGGATATCCA	GCCTTCTTCT	TATCTTTCGA	AGGGGACAAA	CGAAAAGTTG	960
TCGGGAAGTT	TTTACCAACA	GCTTTGAAAA	ATGCGTCTGT	TAATGCGCTT	TTAGTTGTAG	1020
CGAATATCAT	TGCTGTTTAT	CTGATTGGAC	AAAATCAAGG	ATTTTCTTCT	TTAGACACCA	1080
CGACGTTAAT	GTATTATTTA	TTAGTTGGAA	TTAGCTGTAT	GGCTGTTGTT	CGAGCATGTC	1140
TGCCTTTAAA	TCCACTACGC	ATTTTCTTGG	TTTTCAGTAC	CATTATTGGT	ATTTATGTAG	1200
CAGCTATGTT	ATTCCACAAC	ATTTTAGAAA	TTGGTTTCTT	AACTTCGCAA	ACAATGGGTC	1260
TCTTCTTTAT	TATGATGGCC	ATTAATATTG	TGGTGCGTGT	AACAATTGGT	TTCGTACAAA	1320
TGAAACGTGC	TGGA AAAACA	ATTAAAGATT	TATAAAAAGC	AGAGGTTGAG	CTAAAAGCAT	1380
TCGTCTCCAG	AAATGCAGTT	CTTCTTTAAA	TCTTGAAGGA	AGTCGGAGCA	AGCTTGATGT	1440
TTAACTGATT	ATTTTAAGAG	TTAGTGTCTG	GGGCAGGAGT	CAGGAAGACT	TTTGTCTAG	1500
ATACTTTCTT	TTTTTGTGTA	GTTGAAACAA	GATACTTGAA	AAACAGCGAT	TATGTGATAC	1560
AATAAATCAT	GTGTAAATGT	GTCCATTGGA	ATTCGCACTG	AAATGACAGT	GCGTTTTTTA	1620
AACGGTCAAT	CACATGAGTT	AGCACAAATGC	TATTCTTAGG	AGGTGAGGAC	AATGGCACAA	1680
CGCATTCCCTC	AGGAAGTCAT	TGAGGAAGTT	CGTCACCGAA	CGAATATAGT	AGATATTATT	1740
GGGCAGTATG	TCCAGTTAAA	AAAATCCGGT	AAAAATTATA	TGGGTTTATG	TCCGTTTCAC	1800
GAAGAACGTT	CCCCTTCATT	TTCAGTAGCA	GAAGACAAAC	AAATTTTTCa	CTGTTTGGC	1860
TGCGGGAAAG	GTGGAAGTGT	TTTTAATTTC	TTGCAAGAAA	TTGAAGGCAT	TAGTTTCCA	1920
GAATCCGTGA	AACGTGTTGC	AGATTTGGAA	CATTTATCTG	TGGACTTTGA	TTGGTCAGAG	1980
CCGCGTGAAG	TAGCGGATAC	CCCAGAAAAC	CAACAAAGAC	GGAGTTTGTT	ACAACTGCAT	2040
AGTAAAGCCG	CAGAACTTTA	CCATCATATT	TTAGTGAATA	CTAAAATTGG	CGAACCCGCA	2100
TTAACTATT	TATTAGAGCG	GGGGTTAACA	CAAGAACTAA	TTGAGACCTT	TCAAATTGGA	2160
TTTGcGCCGC	AAAAAcGTGa	TTTCTTAAGT	CAAGTTTTTA	AAAACGAACA	ACTGGATGAA	2220
ACGCTTTTTG	AGCCTTCTGG	GCTGTTTGTC	CAAAGAGATA	ACGGTACgTT	TTTGGaTCGT	2280
TTCTATCAAC	GAATCATGTT	TCCAATTAAT	GATCCCCAAG	GCAATGTCAT	CGCCTTTTCT	2340
GGTCGCTTGT	TaAAAACAGC	TGATTTTCCa	GGGGATGAGA	tGCCTAAATA	CTTGAATAGT	2400
CCTGAAACa	CACTTTTTTA	TAAGCGAGAA	ACGCTGTTTA	ACTTTGATAG	AGCCAGAAAA	2460
GAAATTCGTA	AAGAAAATAC	GGTCTTGCTT	TTTGAAGGAT	TCATGGATGT	TATCGCTGCT	2520
TGGCAATCAG	GCGTAAAAAG	TGGGGTGGCT	TCAATGGGGA	CTAGCTTGAC	CAATGAACAA	2580
ATTCGGCGTT	TGGAACGGGT	CGCTAAGGAA	GTAGTTATTT	GTTACGATGG	CGATAATGCC	2640
GGTGTTCaAG	CCACAAACCG	TGCGATTCAA	TTATTGCAAG	AAAACAGTCA	TTTTGACTTG	2700
AGTATTGTCA	GCATCCCTGA	AAAGCTGGAT	CCTGATGAAT	ATGTGCGGAA	ATATGGCGCA	2760

GAAGCTTTTC	AAAATTTAGC	CAATCATGGC	CGAGAAACAG	TTTTTAGTTT	TAAAATGAAT	2820
TATCATCGGT	TAAC TAGAAA	CATGAATAAC	GAAAAAGAAC	AATTGGATTA	TGTGAACGAA	2880
TTGTTACGGG	AGTTAACCAA	TGTTCAATCT	CCCTTGGAAC	GTGATCGTTA	TTTAAACCAA	2940
ATTGCTCAGG	AATTTCAACT	TTCTGTTTCA	AGTTTGGAAG	AGCAATTCAA	TCAATTGAAG	3000
CAAGAGCAAC	GTTCAAGTCA	ACGGCAAGAA	AGGCAACAAT	TTTATCAAGA	TGAAATGATG	3060
CCACCACCAA	TGGAAGAACC	TGTTTTTGAA	GAGAATCACG	TCCAGAACAA	GTTACCGTTA	3120
ACCCAAGTTC	AAAAAGCAGA	ACGTTCTTTA	TTATTTTCGT	TAATGAATGA	ACAAGGAGTA	3180
CGGCAAACCA	TTCAACAATT	GCCAGATTTT	TCATTCGCTC	ATGATGAATA	CCAAGAATTA	3240
TATTTCTTGC	TAGAAAGTTA	TGCGACACTG	CATCAAAGTT	TCGACATTGC	CGATTTTATT	3300
AATTTCTTGC	AAGATAATCA	AACCAACAA	TTAGCAATTG	AAATTGCGTA	TCAGAATTTA	3360
TCGGAAGAAA	GCTCTGAGCG	AGAAGTTGCC	GATTTATTAC	ATGTAATTGC	CTTATCAAGT	3420
ATTGCAGAGG	CAATTGAGCA	GAAAAAGATT	CAACAGCAAG	AAGCAAAACG	AGTGGGCAAC	3480
CAACAGCTTG	AAGCCGAATT	AACAATGGAG	ATTATCCAAT	TGGCTCGACA	GCTTAAAGCT	3540
CAACGAACAT	TTACTTAAGC	AATATAATGA	ATAGAACTAA	GCGTTTTGTT	CATTTTATTT	3600
TGATAAGTCC	CAGTGATTTA	TCAGTTATTT	CAAGCGTGAA	CCGAAAATG	AAGGGGGCCT	3660
TCTTTCATGG	AAAAAGAAAC	AAGTAAAAAA	TATGAAGCTG	CAGTGGCAGC	ATTTATCAAA	3720
GAAAACAAGC	CTAAAGGAAC	AGTGGTTTAT	GATGATTTAG	CGAATCAATT	AGCCACACCG	3780
TATACTTTGG	ATGCAGPAAC	TATGGA AAAA	TTGATCCaAA	AAGTTGAAGA	TGCTGGTATC	3840
AGCGTGGTCG	ATGAAAATGG	AGACCCAAGT	GAACATAGCT	TGAAAAAAGA	TGAAAAAGAA	3900
GCGGAAAAGG	CCCAAGCAGA	AGACTTATCT	GCGCCAACAG	GTGTTAAAAT	TAATGATCCC	3960
GTGCGTATGT	ACCTAAAAGA	AATTGGGCGC	GTTCAATTAT	TAAGTCCGC	AGAAAGAGTT	4020
GAATTGGCAC	TTAAAATTGA	AGAAGGCGAT	CAAGAAGCAA	AACAACGTTT	AGCAGAAGCT	4080
AACTTACGTT	TGGTTGTTTC	AATTGCCAAA	CGATATGTAG	GTCGTGGTAT	GCAaTTCTTG	4140
GATTTAATTC	AAGAAGGTAA	TATGGGCTTA	ATGAAAGCTG	TTGAAAAATT	TGACTACCGT	4200
AAAGGATTCA	AATTCTCTAC	GTATGCTACT	TGGTGGATTG	GTCAAGCGAT	TACGCGGGCA	4260
ATTGCGGACC	AAGCTAGAAC	GATTCGAATT	CCTGTTTACA	TGGTTGAAAC	AATCAATAAA	4320
TTGATTGCGA	TTCAACGCCA	ACTATTGCAA	GATTTAGGTA	GAGAACCAAC	GCCAGAAGAA	4380
ATTGGTGCGG	AAATGGATTT	ACCAACAGAA	AAAGTTTCGT	AAATCCTAAA	AATCGCACAA	4440
GAGCCAGTCT	CTTTAGAAAC	ACCAATTGGT	GAAGAAGATG	ATTCACATTT	AGGTGATTTT	4500
ATTGAAGACC	AAGATGCTAC	CAGTCCTGCT	GAACATGCAG	CTTACGAATT	GTTAAAAGAA	4560
CAACTAGAAG	ATGTTCTGGA	TACTTTAACA	GACCGTGAAG	AAAATGTTTT	ACGTTTACGT	4620
TTTGGTTTAG	ATGATGGTCG	GACACGTACG	CTAGAAGAAG	TCGGAAAAGT	TTTCGGTGTC	4680
ACTCGCGAAC	GTATTCGTCA	AATTGAAGCA	AAAGCGTTAA	GAAAAC TAAG	ACATCCTTCT	4740

CGTTCAAAAC	AATTAAGA	CTTTTAGAA	TAAACATACC	GTTTAGgGGC	TCTTTTCGTG	4800
CATGCGAGAA	GAGCCGCTTT	TGATTTTAA	AAAGTCTCTA	AATGTCATTG	CTAGTTATTG	4860
GAAGAAGGCG	CAAGTTGTGG	TAAATTAGAA	GTAATGGAAT	ATTAGCAGAA	TGTGAGGCGG	4920
AAGAATGACA	AAAAAATGTC	CGAAGTGCGG	GAACGAATTC	GATGCAGAAT	TAACGACCTG	4980
CCCAACATGT	GGCTATTAC	TAACTGACAC	AACAGTTGAT	AAAGAAGAAG	CGGAAACGAC	5040
TTCAACAAAT	ATAGATTTTG	AAACACAGGA	AAATGAAGAA	CACGAAGATC	AATTAAATGA	5100
GAATATTGAG	TGGTCTGAGT	TAAAAGATAT	GAGCTTAGGC	CATGTGATGG	AACTGTTTGG	5160
CGAATCACCA	GAAGAAGAAA	GTAATGATGA	CAAAAAAGAA	GAGTCGACAG	AAGATAACTT	5220
AATAGTCTCT	GATTGAGAGG	ATGTTTCTGG	TCTAGAAGCA	TCGCTTCAAG	AAGGGGCTAG	5280
TGAGGAGACT	CATGATTCGG	TGGAGGAGTC	GATACCAGCG	ACAGAAGAGA	CGCCTACTCA	5340
TTCAACAGAA	GAAAAACTAG	CAACGGACGA	AGCAGTTAAT	TTAGAGGAAA	CAACAGAAGA	5400
AACAACGGAA	GAAACAATA	CTGTTGAAGC	AGAAACAGCT	GAGGTATCTG	AGACGGTGAA	5460
GTCAGAAGAG	GAAGCTTTAA	CAGAGATTTT	AGGAACTGAA	GTTATCTCTA	CCACTTCTGA	5520
AGAAGAAATT	TTTTCACAAC	CACCGATAGA	AGATCAAGAT	GTAACACCAA	ACGAAACTTT	5580
GCAAGCTTAT	ATTCAAGCGC	ATCGAGCAGA	TACAGAAATG	TCTGAAAATC	CATCAGAAGA	5640
AACAGCAGAG	ACCCAAGAGT	TAGAAAACAG	CGGTGAGGCA	GTATTAACCC	AAGCCGAAAC	5700
ACCAACTGAA	TCCATTTTCT	ATTCTGAAGA	AGGGTTAACA	AGTGCAGCGT	CGCTAGAGGC	5760
ACCGAGCGAC	GCTGATGCAA	CCGAGGGAAA	AGTTTCAGAT	TCAGGTGTGA	TACCACCGAT	5820
GAATGAACT	GGGAATGCTC	AACCAGCGCC	TGCTCCGAAG	AAACCATCAA	AAAAAGTAGT	5880
TTTCGTTGCG	CTTGCGGTTG	TTTTATTAGC	TGGAGGTAGT	GCTTGGGCGT	ATCATGATCA	5940
AACGCAAAAA	GCTGCCGCGC	AAGAGGCGGC	GGCACTAACC	AAGAAAACAG	ATACGCTGAA	6000
AGATGAACATA	GCAGCCTTTT	ATACCACGAA	GGAGCAAGTT	TTTATTAAAC	CCGATATGGT	6060
GACTGTTAGT	CCAGAAAAAC	TCTCAAAACA	AGTCGCGGAA	ATTAAAGATT	CAGAAGAGTA	6120
TTCTCAATTA	AATAAACAAA	TTCAAACATT	GAAAGAGAAG	CAACAAACCA	TTCAACAAAT	6180
CAACCAATTA	TTTGAGGCGC	CGATTGTCAA	TGGCAATGAG	TTGAAACCGG	CGATTCTAGC	6240
TGCTGATCAA	CCAATTTCCG	TAAAGAAATT	AACAGGAAAT	GATCCATTTG	ACCAATTGAT	6300
GAACCAAGCA	ATCGATCAaG	CGAACCaaCA	ATACAACCAA	TTACAAAAAG	CTAAAAAAGC	6360
GGTTGAGGTC	ATTTATAAAG	ATGGAAAAAC	TACCAATCAG	CTGAATCGAG	ACACGTATCA	6420
AGCTGCTAAG	GCAGAAGTAG	ATAAAGTAAC	AAGTGATAAG	CTGAAAAAAG	AACTGGTCAA	6480
ACAAGTCACG	ACAGCTGACC	AAGCGTTAAC	CAAGGTTGAA	GAAGAACAGA	AAAGAATTGC	6540
AGAAGAACAA	GCTGCCGCAG	AGCAAGCAAA	ACAAGCAGAA	GAACAAGCGA	AACAAGCTGc	6600
TGCCGCAAAA	AAAGAGGACG	CTAAGAAGGA	AGAAACAGCA	AAAACAGAAG	CGAATGGCTA	6660
CACGGCGCCG	AATAGTGACG	GGGTTTACAC	TAGTCCGCTT	TATGCGCCAG	ATGCTGCCGA	6720

TATCGCTGAT AGTAGCAATC CAGCGTGGAC ATGGGCACCA GGTGTGAAAG AAAAAGTCTT	6780
AGACACAGTG ATTGCTCGTG GCTATGTGGT TCCTGGCGGA TATTCTTTAG AACCTGCCAA	6840
AATTGTGAAT GGCGAAGGTT ACTACAATCT TTATGCAACA AACAATCAAT CAAAATTATT	6900
AGAAGGCACT ACAGAGAAAA ATGTTTCACAT GTATTTAGTT ACAATCAATG CTAAAACAGG	6960
TTGGTTCAAA GGGAACGCTT CTCGTAATGC AGGGCAATAA AAGAAAATAG AAACAATCGA	7020
ACAAAAGCGG AAATCTTTTG TTCGATTGAT TTTCTTGAG GAGTTAATGA TAAATGAGAG	7080
AATTACTAGC AACAACATTT ACCGGCATTG TGGTCGATGA AAATGAACAA TTTTATTTTG	7140
TTCAAAAAAA TGGCATCACA TTCCGTTTAA AAAAAGAAGA AGGTACGCAT GCTATAGGGG	7200
AGGCTGTAGA AGGGTTTGGT TATTTAAATC AAAACAAGA GCCAGCACTG ACAACTACAA	7260
TTCTACTGT GCGCATTGGT AGCTATGGCT TTGGTACTGT CACAGGAACG AGACGTGATT	7320
TAGGTGTCTT TGTAGACGTT GGCTTACCAG ATAAAGATGT CGTTATTTTCG TTGGATGAAT	7380
TACCAACTAT GCGGGAACCT TGGCCTAAAA AAGAAGACCG AGTAATGGTT ACTTTAAAAG	7440
TGGACAGTAA AGACCGCATT TGGGGCGAaT TAGCAGnTGA nAAAGTCTTT AAAGCAATGG	7500
CAAAACGTGG GAATGCTGAA ATGCAAAACC AAAACTTAAC CGGGCATGTG TATCGTCTGA	7560
AACTTGCAGG AACGTATATT TTAAGTGATG ATTTATACAT TGGATTTATC CATCCTTCTG	7620
AACGATTCCA GGAACCACGT TTAGGTGAAA AAGTAGTGGG GCGTGTGATT GGCGTGCGGC	7680
CGGACGGCGT ATTGAATCTT TCATTAAAAC CAAGAAGTCA TGAAGTAATC AGCGATGATG	7740
CGTTAATGAT TCTAACTTTC TTAGAAAGAG CAAACGATCA CCAAATTCCT TTCACCGATA	7800
AATCAAGCCC TGATGAAATT AAACAAACCT TTGGCATTAG TAAAGCTCAA TTAAACGCG	7860
CCTTAGGCCA TTTAATGaAA CAAAAATTAA TTAAACAAGA AgATGGTAAA aCAATTCTAA	7920
TTGGcTCGTT GGrACAATCm AAAGAAA	7947

## (2) INFORMATION FOR SEQ ID NO: 85:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7865 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

GCACCAGCGT GGTTGGATCA TTCTGATCAC CCGTTTCATT GGCCCGATAC TGGTGTTTAT	60
TTATATGAAC CTGGCCAAGA AGTTGTTTTT CAATTCGATC AATTGGTAAC AGATATTCGT	120
GAATGACCCA GAAACCAGTT AAAAAGCAAT TGCCTTTAGG GTTAACTACT TTATTTGCTG	180
TACAATTtAT GmAAGAAaTT TCACGATTTT TAaCAACCCA TCCACATGTG AaTTTAATAT	240
TACAGCArGA CGGCTCGCCA AAaTTACAAa CGATGTTAGC AAATAAAGAA ATTGATATGG	300
GACTAATTTT TTTCCCAAAT ACCTTACCTG AAATTATTCA TATTGAACCT TTAGAAACGA	360



CTACCAAAGG	CTACCATGTT	TATGTCGTAG	TTCCAGAATC	AAATCCTCTC	TCCCAATATG	420
AAAAATTAAC	TTTAAAGAT	TTGAAAGATC	AACgATTTC	ATCGTTAAGT	GATAATTTTA	480
TGATTGGTCG	CCTACTACTA	GATCGGACTC	GGAGCTTCGG	TTATGAGCCG	AATATCATT	540
TGCATAACGA	TGATTTACAA	GTACTTCTTT	ATAGTTTGCA	AAAAATAAT	TCGATTGT	600
TGCTGCCGAT	TGAATATTAT	GAAGTGGGAA	AAAGTCAGGG	ACTGAAATGG	ATTCCtTTAA	660
AAGATAAGTT	TGACTATTTT	CCCaTCGGCA	TTGCTTTGCG	CCGCGATTTT	AGTCTGACAG	720
AAGATGTTAG	AGATTTTATC	CaAATCATT	AAGAAAATTA	ATGGATAATC	TAGCCTTGAT	780
AAGCAAAGAC	GATTTTACTA	TCATAAATTT	TCAACACTTT	TTCAAATAAG	AATTTTGAGA	840
ATAAACCCGC	GACAATTGGA	ACAACAAACC	AAGAAAAGAG	TAAGGGTAAG	ATCCCTAAGC	900
CCGCATCCAA	TGAAGCAAGT	GGTCCACGA	GTCCAACCAA	GCCAAATCCA	GCCGATTGTG	960
GGGTTCTCTGA	AATGTTGAAT	AGAACCACAG	GAATTGCTGA	AATAGTTGCT	GTAAATAAAC	1020
AAGGAATTAA	AATAATTGGG	TATTTAAATA	AGTTAGGCAT	CATCATTTTC	ATGCCGCCCA	1080
AAGAGACAGC	CAGTGTCACT	CCCgATTGGT	TGACATTCCA	AGAATTAATC	ACAAGAACAA	1140
CAGTTGTGGC	TGCAATTCCC	ATTGCAGCAG	CACCTGCAGA	CAAGCCATTC	AACTGAATCG	1200
CCAAACCAAT	TGCCACAGTA	GTGATTGGCG	AAATAATCAG	TGCAGCAAAT	GAACAAGCAA	1260
TTAAATACT	CATGATAATT	GGCTGAAAAT	CAGTAAAT	GTTAATCACT	TTCCCGATGG	1320
CAACCGTAAT	TTGTGTCACA	TATGGATAAA	TCAACATGCC	AAACAAACCT	GCACCTACTC	1380
CAACAACGAT	CGGTAAAGCA	ATGATTTCCA	CTGAACCGAA	TTTTTTATCG	ATGACTAATA	1440
AAAGTAAGAC	TGCAACAGAA	GCGGTAATCA	TAATATTAAT	GATGTCACCA	GTGCCTGCAC	1500
CAATAAAGCC	TTTCACtTCT	GaATTAAAT	TAATCACGCC	AGAGCCAGCA	AATGCTGCAC	1560
CGCCAGCAAT	CATCATTTTC	TGTGGCGTTA	AGCCAAATG	AAAAGCmATC	AGTCCACCAA	1620
TAATTAAAGG	GGTGGCTAGT	TGGAAAATTT	GCGCCGCATG	AATAATCATT	TCGATAATCT	1680
TGTATTCTGC	AAAATATTTT	AAAATGGCTC	CTAATACAGC	ATTGGGAATT	AAGGCGATAA	1740
TGGTGCCGGA	AgcTGTTCCA	GCTAACACTT	TGTTAAAAA	TATTCTTGGG	GTTAATTGAG	1800
CATCCAGTTT	ATCAGTGGGC	ATTTTCAACA	CTCCTATTGA	TTATTTTGT	TTTTTTAGAA	1860
CAAGTTTGAG	AATGATTGAA	CTTTCTATTA	GATAAAGAGC	GGCCTCAAAG	AAGAGCGCTT	1920
ATTTCTTTGA	GGCGCTTTTT	TTTTAACTAA	AGCAAGGAAC	GATTTTAAAT	ATATATCCGT	1980
ATCCGCATTA	TATTTGCTA	AATTATCTTT	AACATCTCGc	AAGAACTTAC	CTGcTTGTTG	2040
TCCATCTAAA	ATTGATGAT	CAATAGAAAG	ACATAAATTA	ACCATATCTG	CCACTTTAAA	2100
GCCTCCATCT	GCGGTTGGCA	CAAGGCGTTT	ATTAATTGAC	TCTACTTGTA	ATATCGCAGC	2160
TTGTGGATGG	TTAATAATTC	CCATCGATTG	GACGGAACCT	AGTGTTCAG	TATTATTAAG	2220
AGTAAAGGTC	CCTCCTTGCA	TTTCTTTGCT	TGCTAGCGTT	CCTTGTCGGA	CTTCTTGTC	2280
CAAGCGATTG	ATTTCTTTGG	CTAGCCCTGC	AATCGAATAG	TTATCCGCTT	GTTGAATCAC	2340

586

TGGCACATAA AGATGCTCGT CTGTCGTTAC AGCTATGGAT AAATTAACAT CTTTATGATA	2400
AATAATTGAG CCATCATCCC AAGATGTATT GATTTTGGGA TTTTCTTTA ACGCTTGAAT	2460
GACTGCTTTG GCAAAAAATG GGAAGAAGCT TAGAATAGG CCCTCTTGTT GTTTAAATTC	2520
ATCTTTTAAT GAATTTCTAA GTTGGACTAA GTTGGTCACA TCCGCTTCCA CCATTAACCA	2580
AGCGTGAGGG ATTCGTTGA CGCTTTGGAC CATTTTTTTA GCAATTGCTT TACGTACAGG	2640
ATCTGCAGAG ACAATTTTAT CTGTGCTTGT TTCAATGGGA CTAGCTGTTG CGGCGCTTTC	2700
ATTTTGCGAA gCAACTGGCT CCTCTGAAT TGAGGGGCTA GTTCCTGGAC TCACTGTTTT	2760
CTCAGGTGTT CTTGCCTGTG TTGGTGTAAA GTTTGTAAACA TCTTTTCTAG TAATTCGACC	2820
ATCACGACCT GTCCCTGTAA CTTGTGTAA ATCAATCTTT TTTTCTTGGG CAATTTTAAG	2880
GACTGCTGGG GAATAACGAC CATTATTTTT TTGATGGGAC GTCGCAGTGC TTGTTGTTTC	2940
AGCCGTTTCA TGTTCTTGGC CTTGTCCGC ACTTGCTTCT TTTCACTGGT CTAAAGTAGC	3000
TACTTCTGTT TCTTCTGTTG TTTCTTCTGT CTCTACGTC ATCACCGCTG TTCCGATGGG	3060
CACATCTGTA TCTAGAGAAA TCAGAAATTC TTTTACTACA CCATCAAAAT CTGATGGGAC	3120
TTCCGTGGTG ACTTTATCAG AACGACTTC CATTAAAGGA TCATAGCGTT TCACTGAATC	3180
TCCTGGTTTA ACTAACCATT GGACAATGGC TGCCTCTGTG ACACTTTCGC CTAGATGAGG	3240
CATTTTGATT TCTTTTGTGG CCATTATTTT TTTCTCTCTT GCTTCAATTA AAATTCTGCT	3300
AATTCTTTCA TTGCTGCTAA TACTTGCTCT TCATTGATTA AAAATTCTCT TTCTAAAGGT	3360
AATGCATAAG GCATGCTTGG ACAATCAGGT CCTGCTAGTC GTTGAATCGG TGCATCTAAA	3420
TCAAAGAGAG CATCTTCTGA AATCAATGCC GCAATTCAC TCATCACGCT GCCTTCTTTA	3480
TTGTCTTCTG TTACGAGTAA GACCTTCCCT GTTTTCTTAG CTGCAGCCAC TAATGTTTCG	3540
CGATCTAAAG GATATAATGA GCGAACATCG ACGATTCGG CATCAATTCC TTCAGCGGCT	3600
AGTTTTTCCG CCGCAGCTAA AGCTAATTGC AAGGTCATGC CGTAACTAAT CACGGTTAAA	3660
TCACTACCCG TTCTAACGAC APTGGCTTTG TCAATGGTA CGATATAATC ATCTGCAGGC	3720
ACTTCATCTT TTAGTAAACG ATATAAACGC TTATGTTCTG AAAAAATGAC TGGATCATCG	3780
GAACGAATCG CTGCTTTAAT CATCCCTTTT GCATCATAAG GATTAGAAGG GGTCACTACT	3840
CTCAATCCTG GTTGTCCACA AAAGACTTTT TCTGTAGACT GAGAATGATA CAGCCCACCA	3900
CGAACACCGC CGCCATAAGG GGTCCGATAA ACCATTGGCG CAGTCCAATC GCCTTTTGTT	3960
CTATAGCGCA TTGTCCGAGC CTCTGATAAC AATTGATTGG TTGCAGGCAA AATGTAATCC	4020
GCAAATTGAA ATTCACCAAT TGCGCGATAG CCCATTAAAC CTAAGCCGAC AGCTAAACCA	4080
CCAATTAAGC CTTCTGTAA CGGTGTATTA AAACAACGCT CGTCACCGTA TTTAGCAGCC	4140
AAGCCCTTTG TTACACCGAA CACGCCGCCT TTGTCCCGC CGACATCTTC TCCAAAAATA	4200
ACTACTTTTT CATCACGAGC CATTTCTTCA GAAATTCCTA AGTTAATTGC TTCTAAATAA	4260
GTCATCTCAG CCATTATATT TCTTCTCTC TTCTCTTATT TTGCATACAC TTCTTCTAAA	4320

ATTGATGTCG GTACAGGATC TGGCATTGCT TCTGCTTCAT CCGTTGCTTG ATTGATTCT	4380
GCACGAATTT CTTTCATCAAT TTTGGCAATG TCTTCGTCTG TTAAATAGCC CTCTTCTAAT	4440
AATTGTTTTT CAAAAAGCTT CACTGCATCG TTCTTTTTCA TTTCTTCAAT TTCTTCTTTT	4500
GAACGATAAA CAGATTGATC ATCGTCAGCG GAATGAGACG TCAAGCGCGA AACCATTAAT	4560
TCAATCAATT TTGGTCCTTT TTTCCCGCGA GCCGCTTTTA CTGCTTCTTT AAATGCTAGA	4620
TAGACTTCAG TAAATCACT ACCATCAACG GTTACACCTT CAAAGCCATA AGCTTTCGCG	4680
CGATCGGCCA TTCGTTTATT GGCATACTGT TCTTCAATTG GGACAGAAAT CGCATATTCA	4740
TTATTTTCAA CAACAAAAAT AACTGGTAAT TTTTTTACGC CTGCAAAGTT CATAGCTTCT	4800
TGGACTTCTC CTTGATTGGC AGAGCCTTCC CCAGTGGTGG TTAATGCAAC AAAATCAGCT	4860
TTTTGAAGTT GCGCTGCATA ACCAACACCT GTCGCTAATG GCATTGTGT ACTTACTGTT	4920
GAAGAGAAGG AAACAATATT ATGCTCTTTT GAACCATAAT GATTCGGCAT TTGACGACCA	4980
TGGGAAGAAG GATCCGCTTC TTTTCCAAAA GAACCCATTA AAATATCTTT GGAGGTCATG	5040
CCCCAAACCA AGCACGCGGT CATATCACGA TAATACGGTA AAAATAATC TTTTGGAGGA	5100
TCAAAGCCA TCGCCATCGC TACTTGTGCA ACTTCTGCCC CTTGACCAGA AATATTGAAA	5160
GAGGTCTTAC CAATCCTTGT TAATTGCCAC AACCGTTCGT CTAAACGaCG tCCTCTTAGT	5220
ACTTcAcGAt ATGCTGcAT CAaTtCTTCT TTTGATAAAC CTGATTTTTT AAGCGCTTTC	5280
ATTTTCATCA ACCTTTCTCT ATTTGTGAAT CGCTAAGCCA TAAGTATCTA AGGCTGCTTC	5340
TTGCAATACT TCTGTCATTG TTGGATGCGC ATGGATTGCT TCACCAATTT CAATGGGCGC	5400
GGCATCTaAA TACATGGCAG TACTTGCTTC GGCAATTAAA TCCGTGACAT GTGGGCCAAT	5460
CATAGAAACC CCTAATAAAT CATCGGTCTT CTTGTACGA ATTACTTCTA TAAAGCCATC	5520
TGTTTCTCCA TAAACAAGTG ATTTACCATT ACCATTAAAA TTAAAGGTGC CAATTACTAC	5580
TTCTTTTTCT GCTGGTAAAG TTTCTCTAGT ATAGCCGACA CTTGCTATTT CAGGATTTGT	5640
ATAAACACCT CGAGGTACGT TTGTATAATT TAAAGGTTCA ACTGTCTCGC CTAAAAGATG	5700
CTGAACAGCT AACTCTCCTT CTTTCATAGC CACATGTGCT AGTTGGAGTG TATCGATACA	5760
ATCACCAATC GCATAAATAT GTCCTtCTGT GGTtTGATAA AATTCATTTA CTTCAATGCC	5820
TTTGTcAGTA TATTTCACTG AGGTATTCTG TAACCCTAAC TTATTGATAT TGGGCTGACG	5880
TCCAATTGCT ACCATGACTT TGTCAACGGT GAGGGTTTCT TGTcCTGCAA CTTCTACTTG	5940
AACCTTTTGT CCAGTAACTT TCGCTTCTTG GACTTTGCTT CCTAATAAAA TGTTAATTCC	6000
TCGTTGCTCT AAACGTTTCT TTAATTCTTT AGAAATTGTT GCACTCTCAT TTATAAGTAA	6060
ACGATCCAAA AATTCAATGA TAGTTACGTT GACACCTAAA CTATTTAATA AAGAAGCCCA	6120
CTCGACACCA ATAACGCCAC CACCAATAAT TGCAATTGAT TCTGGTAGCT CTTCTAATTC	6180
AAGCATCCCA TCAGAAGATA AGATAAATTC TTCATCCAAT GGCAAATTAG GTAACGTTTT	6240
AGAACTAGAG CCTGTCGCAA TAATGACATT TTTAGGAcGA TAATTTCTtC TTCTCGTGTT	6300

GGAnCATTAA ACGTAACTGC CACTGCGCCA GAGACTGGTG AAAAAATGGA GGGACCTAAA 6360  
 ATCGCACCTT CGCCTGCTAA AATTgATTyT ATTTTCTTG CATAAGCCTT CAACACCTTT 6420  
 GTGTAAGTGC TCAATAATTC CTTCTTTCCG TTGTTGTATT TTAGAAAAAT CAATGGACGC 6480  
 TGCCTCTGTT TCAATCCCAA AAGAAGCCGC TTGTTTAAAG GTATCAAAAA CCTCTGCGCT 6540  
 TCTTAATAAA GCTTTTGTAG GGATACAGCC TTTGTGCAA CAAGTGCCAC CTAATTTGTA 6600  
 TTTTTCACAA ATTGTGACGT TTAGCCCTTT TTGTGCAGCT CGAATTGCTG CTACATAGCC 6660  
 TCCTGTTCCG CCACCTAAAA TCAGTAAATC TGTTTGTTC GCCATTTTGA TCACTCCTGC 6720  
 TCAATTTTTG CTTCACTATA ATTTAAAGCT TGTTCTTCCC CTGTTAATAC TCGATTGACG 6780  
 CCTTCATATA AAGCAGCCAT CTCCATTTCA CCTGGGTAAA CCTTAATTGG TGCAATCCAA 6840  
 GTGACTTTTT GACTAATTTT TTGAACAACG GTTTGCGAAT AgcTGCGCC CCTGTAAAA 6900  
 TAATTGCATC AATTGTGCCT TCAAGAACCA CGGCCATCTC GCCAATACTT TyyGCGATTT 6960  
 GGTAACACAT TCCTTTTAA TAGTAATTTG CTGTTTGATC ACCTGCAGCT ATTTGTGCTT 7020  
 GGATATGCCG TAAATCTGTT TCACCTAGGT ATGATTTAAG ACCGCTATTC CCCGCAATTA 7080  
 GCTTTTTTAC TTGGCTAATC GTTAACTCCT GTTCAAGTAT CCATTGCGCA AATTCAACGA 7140  
 GCGGTAAGGC ACCACTGCGT TCTGGCGTAT AAGGaCtTCA CCATCAgCCC ATTTACCACA 7200  
 TCCACawGCG nCcTTTTgAT GCGCCCCCnA GCTAATGCC CCACCTAAAT GAACAACGAT 7260  
 AAAATTGCTT TGCTCATATG TTTTCCCTAA ATCCTCAGCG ATTTTTCGAG CGACCGCTTT 7320  
 TTGATTTAAG GCGTGACCcA AaCkaCGCCG CTGAATGCCT TTTAATCCAG AAATCCTGGC 7380  
 TAATGGCTGT AGCTCATCTA CAACAACAGG ATCGACAATA AAGGCTGGCA CGTGATATTT 7440  
 TTCTGCAAAC TCATTAGCTA AAATCGCCCC TAAATTTGAA GCATGGGTAT TAAAGCGTTC 7500  
 TGTCCGCAAG TCTTCTAGCA TTTGTTGGTC GACAAGATAC GTACCGCCAG GAATAGGCTT 7560  
 AAGCAAACCA CCACGCCCTA CAACAGCGGC TAATTGTATG ATGTTATGTG TTTCTAAAA 7620  
 TTCCGCGATC ATTTGTTTTT GAAATGATGT TTGACTGACA ACATTTTCAA AAGGTGCCAA 7680  
 TTCTTGACG CTATGTCTAA GCGTTTCTTC TGCCAAGCAA TCATGATTAG CAAAAGCGC 7740  
 TAACTTCGTG GAGGTCGAGC CAGGATTAAT AACCAATACT GTTCCATTG TtTTCTGCTC 7800  
 CTTkTTTTtAA ACTTGTCTCA TGGCAAATCT TAATGAATGG AATTGCTTT CTGTCGAATC 7860  
 GCTTC 7865

## (2) INFORMATION FOR SEQ ID NO: 86:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4023 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TGCTCCnGnT GTGCTCGATG TATTACAAGA AAGATTCCAA ATTCTTCGCA ATATTTATTG 60

GATGCAACCT ATTGGACGTA GAAGTTTATC AGAAACGATG GGAATTACAG AGCGAGTCTT	120
GCGTACAGAA ACAGATGTTT TAAAGCAACT CAACCTGATT GAACCTTCAA AAAGCGGGAT	180
GACATTAACA GAACGCGGTT TGGAGTCTA CCAAGGCTTA GAATTAGTCA TGAATCAATT	240
GCTTGGCATG CACCAAATCG AAAAAGAAAT GACACAGTAC TTTGGCATT CACGTTGTAT	300
CGTTGTTGCA GGAGATAGTG ATATCCAAAA GAAAGTTCTT TCGGACTTTG GAGATGTCTT	360
AACCAATACG CTGAATTTAC TCTTACCAA TGGTGAGAAT ACAATTGCGG TTATGGGTGG	420
GACAACAATG GCCATGGTGG CCGAAAATAT GGGATCACTA GAAACTGAAA AACGTCATAA	480
TTTGTGTTGTT CCAGCCAGAG GAGGGATTGG TGAAGCAGTT TCTGTCCAAG CCAATTCAAT	540
TAGTGCTGTA ATGGCCAACA AAACGGGTGG TAACTACCGG GCGTTATATG TCCCAGAACA	600
ATTGAGTCGT GAAACGTATA ATAGTTTATT ACAAGAACCC TCGATTCAAG AAGTTCTTAC	660
GTTAATTAGT CATGCCAACT GTGTGGTGCA CAGTATTGGT CGAGCGTTGC ACATGGCGGC	720
ACGCCGAAAA ATGTCTGATG ACGAAATGGT TATGCTCAAA CAAAAAATG CCGTGGCAGA	780
ATCGTTTGGT TATTTCTTTG ATGAAGAAGG AAAAGTGGTC TATAAAATTC CACGAATTGG	840
CCTTCAATTA AAAAATTAC AAGAAATTCC CTATGTCGTC GCAATTGCGG GCGGCAAAAC	900
GAAAGCCAAA GCCATTGCGG CATATATGAA AAATGCACCA AAACAAACGT GGCTCATCAC	960
AGATGAAGCT GCCGCAAACG AGATTTTAAA AGGGGTAACC CTTTAAAAATA AAAAATTTT	1020
GATTTTCATA AGGAGGAAAT CTTAATGAC AGTTAAAGTA GGTATTAATG GATTTGGACG	1080
TATCGGACGC TTAGCTTTCC GTCGTATCCA AGATGTAGAA GGAATCGAAG TAGTTGCAAT	1140
CAACGACTTA ACAGATGCTA AAATGTTAGC ACACTTGTTA AAATATGACA CAACTCAAGG	1200
CCGTTTCAAC GGTACAGTTG AAGTTCACGA AGGTTCAATC AACGTTAACG GCAAAGAAAT	1260
CAAAGTTTTA GCTAACCGTA ACCCTGAAGA ATTACCATGG GCGGAATTAG GCGTAGACAT	1320
CGTTTTAGAA TGTAAGTGGT TCTTTACTTC TAAAGAAGCT GCTGAAAAAC ATTTAACTGC	1380
TGGTGCAAAA CGTGTAGTTA TTTCAGCTCC TGGTGGAAC GATGTACCAA CAATCGTTTA	1440
CAACACAAAC CATGAAACAT TAACTGGAGA AGAACTGTA ATTCAGGCG CTTCTGTAC	1500
TACAACTGC TTAGCTCCAA TGGCTAAAGC TTTACATGAC AACTTTGGTG TTGTTGAAGG	1560
TTAATGACA ACTATCCACG CTTACACAGG TGACCAAATG ACATTAGACG GACCACATCC	1620
TAAAGGCGAC TTCCGTCGTG CGCGCGCTGC TGCTGCAAAC ATCGTACCTA ACTCAACTGG	1680
TGCTGCTAAA GCAATCGGCT TAGTAATCCC AGAATTGAAC GGTAAATTAG ATGGCGCTGC	1740
TCAACGTGTT CCTGTAGCAA CTGGTTCATT AACTGAATTA GTTACTGTAT TAGACAAAGA	1800
AGTTACTGTT GATGAAGTAA ATGCAGTAAT GGAAAAGCT GCTAACGAAT CTTATGGTTA	1860
TAACACAGAC GAAATCGTTT CTTCTGATAT CGTAGGTATG ACTTACGGTT CATTATTCGA	1920
TGCAACTCAA ACAAAGTGA TGACAGTTGG CGACAAACAA TTAGTTAAAA CTGTTGCTTG	1980
GTATGACAAC GAAATGTCAT ACACTGCTCA ATTAGTTCGT ACTTTAGAAT ACTTCGCTAA	2040

CTTATAAGAT TCATTCTTAT CATTAGTGAA ATTTGAACAA TTGTGAAATA ATAAGCGGGG	2100
AAGCAACGCG CTTcCCGCT TTCTTTTTTT AAAAAGAAA AATCCTAAAA TTGACAATAC	2160
AGGAGGTCTA CTCATGGCTA AAAAGACAAT TAAAGATGTA GATTTAAAAG ACAAAAAAGT	2220
CCTTGTCCGT GTTGACTTTA ACGTTCCGTT AAAAGATGGC GTGATCACAA ACGATAATCG	2280
TATCGTAGCG GCTTTACCAA CAATCAAATA CGTGATTGAA AATGGCGGAA AAGCAATTCT	2340
TTTCTCTCAT TTAGGTCGTG TAAAAACAGA AGAAGACAAA GCTGGCAAAT CATTAAAACC	2400
AGTGGCTGAA CGTTTAAGCG AATTACTAGG TCAACCAGTT ACTTTTGTTT CAGAAACACG	2460
TGGTAAGAA TTAGAAGACG CTGtTAACAA TATGAAAGAT GGCGACGTTT TAGTATTTGA	2520
AAACACTCGT TTTGAAGATG TTGACGGCAA AAAAGAAAGC GGCAATGACG CTGAATTGGG	2580
TAAATACTGG GCTTCTTTAG GCGATGTATT TGTCAATGAT GCTTTTGGTA CTGCACACCG	2640
TGCACATGCC TCAAACGTTG GAATTGCTTC AACTGGTATC CCAACAGTTG CTGGTTTCTT	2700
AATGGAAAAA GAAATTAAAT TCATCGGCGA AGCTGTTGAA AATCCAAAAC GTCCATTCGT	2760
AGCAATCTTG GGTGGCGCAA AAGTTTCTGA CAAATCGCT GTCATTGAAA ACCTAATTGA	2820
AAAAGCCGAT AAAATTTTAA TCGGTGGCGG TATGGCTTAC ACATTCATGA AAGCACAAGG	2880
CTATAGTGTT GGTATCTT TACTTGAAGA AGATAAAGTG GACTTAGCGA AAAGCTTAAT	2940
GGAAAAAGCT GGCGACAAAT TAGTTTTACC GGTGATACA GTTGTCTCTA AAGAGTTCAG	3000
CAATGACGCT CCTTCCACA CGGTTCTTC AACTGAAATC CCAGATGATG AAGAAGGCTT	3060
GGATATTGGT GAAAAACAA TTGAATTATT CGCTAACGAA TTACAAGGTG CGAAAACAGT	3120
TGTTTGGAAT GGACCAATGG GCGTATTTGA AATGAGTAAC TTTGCCAAAG GAACAATTGG	3180
TGTTTGTAAG GCGATTGCTA ATTTAGAAGA CGCAACAACA ATCATCGGTG GCGGGGACTC	3240
TGCTGCTGCA GCAATTCAAT TAGGCTACGA AAACAAATTC TCTCACATCT CAACAGGTGG	3300
CGGAGCAAGT TTAGAATTAT TAGAAGGCAA AACTTTACCA GGTTTAGCTT CTATTAACGA	3360
TAAATAATCA GTTTAAAAGA CAATATTTTT AAAAAGAAA AAGGATGTGC TTTTCATGCG	3420
CAAACCAATT ATCGCTGGTA ACTGGAAAAT GAACAAAAC TTAAGCGAAG CACAAAGCTT	3480
TGCAGAAGCA GTAAAAATG CTGTTCCATC AAACGACGTT GTTGACGCTG TGATTGGATC	3540
TCCAGCATTG TTCTTAGCTC CTTTAGCTTG GAACCTAAAA GATTCTGAAG TGAAATTAGC	3600
TGCACAAAAC TGCTACTGGG AAAACGCTGG TGCCTTCACT GGTGAAAAC CACCAGCTGC	3660
AATTGCAGAT TTAGGTGTAG ACTATGTGAT TATCGGACAC TCTGAACGTC GCGAATATTT	3720
CCACGAAACA GATGAAGATA TCAATAAAAA AGCCAAAGCA ATCTTTGCAA ATGGTATGAC	3780
ACCAATCTTC TGTTGTGGTG AAACCTTAGA AACGTATnAA GCTGGTAAAA CAGCGGAATG	3840
GATCGAAGGT CAAATCACTA AAGGCTTAGT TGTTTATCA AACGAACAAG TTGCTTCAAT	3900
GGTTATCGCT TACGAACCAA TTTGGGCAAT CGGAACGGT AAATCTGCAG ATGCTAATAT	3960
TGCAGATGAA ATCTGTGGTG TTGTTCTGTT AACAGTTGAA AAATTATACG GCAAAGAAGT	4020

TTC

4023

## (2) INFORMATION FOR SEQ ID NO: 87:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2288 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

GGGCGGTGTC ACTGCGTCTA TTTTCGGTTA TCGGCCAACC TTTTTCATTA CTGGTACGAT	60
TTTATWACTC GTTTTTGTTT TGAGTCTCGT CTTTGTCCAT GAAGAGTTTG TGCCAATTGA	120
AAAAAATCAA GCCGCATCTG GGAAACAAAT TTTAAAAAAA CTAGAACATC CCCACGTGAT	180
TCTTGGAATG TTTATTACGA CATTAATCAT TCAAGCTTCC AATAATTCAA TTAGTCCCAT	240
CATCAGTTTA TATATTCAAC AATTGTTGGG TGGTCACGGA AATGTCACCT TAATTAGCGG	300
AGTCATTGCT TCTATTCCAG GGATTGCAAC GTTAATAGCC GCCCTCGTT TCGGCCGGTT	360
AGGCGATCGT ATTGGCAGTG AACGCATCTT GACAATTGGC TTAATTTTAG CCATTTTGT	420
TTACCTACCA ATGGCCTTTG TTCAAAATGT CTGGCAGCTT GCCATGCTTC GTTCTTAGT	480
TGGGATTTCC GATGCCTGTT TACTACCAGC TGTTCAAACG TTGATCACTC GTTATTCCCC	540
AAGTGACGCT GCAGGTCGTA TTTTCAGTTA CAATCAATCT TTCCAAGCGA CAGGAAATGT	600
GATTGGTCTT ATGATTGGTT CAAGCGTATC TGCCGCTTTC GGTTATCGTG GGGTCTTTAT	660
TTCCACTTCT TGTCTGGTTC TCCTTAACCT TCTATGGGTT CGTCGAAGCA CCGCCGAATT	720
AAAAAAGAG AAAAATGATG ACTAAGTTAT ACACAGAAAA GTTATCCACA ATTTGAATTT	780
TTAATTTTTT TCTTTTCTT ATTTTAAAAA AAATAGGCAT ATAAAACAAA AAAGAAAAGC	840
GGATATTTAC TGGTTTTATG ACCAGCATAT CCGCTTTTCT TTTGGTAAAC TATACGCTTT	900
TTGTTTCACA ACTCGTGAAA CTTTTTAGAC GTATTGTAGT TCTTTTTCAT CTTGGTAGGC	960
GCGATCAATT AAGCCGCCCC CTAAGCATT C ATCCCATCA TAAAAACAA CAGCTTGACC	1020
TGGAGTGACT GCACGCGCTG GTTCCTTAAA CGTAACGGTT GCTTTTGTGC CATCTTCAGA	1080
TAAGGAAACA TGCACTGGAA TATCTGCTTG ACGATAACGG AATTTAGCCG TACAGTCAAA	1140
CTCTTTTGGC ATTGGTGTAT CAACAGTAAA ATGAACCTCA CTAGCTTCTA AATGTGTTGC	1200
ATATAGTTTT TCATGGTGAA ATCCTTGtCC CACATATAAT GTATTGGTAG TTAAATCTTT	1260
CCCAATCACA AACCATGGTT CTTGTGTTTT ACCGCCGCCA CCAATTCCTA AGCCTTGCGG	1320
TTGACCGATA GTATAATACA TCAGACCATC ATGTTGTCCT TTGATTTGCG CATCTTCTGT	1380
CACCATGTTG CCTTTTTTAG CTGGTAAATA ATTACTCAA AATTCTTTGA AATTCTTTTC	1440
ACCAATAAAA CAAACGCCCC TTGAGTCTTT TTTCTTGGCT GTGGCTAATC CAGCACGTTC	1500
AGCAATTGCA CGAACTTCTG ATTTTTCAT GCCACCCAAT GGAAACATTG TTTTGGCAA	1560

TTGGGCTTGT GATAATTGGC TTAATAAATA GGTTTGGTCT TTATTATTGT CAATTCCACG	1620
TAACATATGC ACTACGCCAT TTTTCATCTGT TTCTACTTGC GCATAGTGGC CAGTkGCAAC	1680
ATAGTCAGCG CCTAGTTGCA TAGCATAATC TAGGAATGCT TTAAATTTAA TTTCTTTATT	1740
ACACATGACA TCTGGATTTG GTGTCCGTCC TAACCGATAT TCTGCTAAAA AATATTCAAA	1800
CACACGGTCC CAATATTCTT TTTCAAAATT GACCGAATAA TACGGAATCC CAATTTGATC	1860
CGCTACTTTT GCGACATCTT TATAATCCTC AGTTGCGGTA CAGACACCAT TTTCTGTCTG	1920
ATCGTCCCAA TTTTTCATAA AGATCCCAAT TACATCATAG CCTTGCTCTT TTAAAGCAA	1980
CGCTGTGACA GATGAATCGA CACCGCCACT CATGCCTACC ACGACACGCG TTTTGCTGTT	2040
ATCTGCCATT TTATCACCAT CATTTC AATT AGATTCTGAA ATCATCCACG ATTTGAAGGT	2100
CGTAACTTTT CAGTGCTTTT CATTATACTA TTTTTTTGAA ATAAGTAAAG CAAGCCGGGC	2160
GTATAATGAG AAAACACTTG GAACAAAGAC ACAGCCAACA AGCTATTGCT TTCAAATAAA	2220
CAGAAAACcT GmAGGAACAA GCTACTTTTCG CTTGTkTCTT CAGGTTTCTC TAAATTATTC	2280
TACTTTCT	2288

## (2) INFORMATION FOR SEQ ID NO: 88:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6691 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

TTTAATTTTT GGGAAaAAAT TwCCaAAGGC TAAGGaTaGk TAAAtTGgGG ACCwTATTTg	60
GAATaTTGaT TAAATTCTGA CCACTTAATT GtCTTTkGAA AAggTTATGC TAAAGTTACT	120
TTGTAATcGA ATGaGGAGGG CTATCACATG GaAGAAAAGA ATTGGCCTGA GTTATTTGCT	180
CGTTATCTTA TTGTCGAAAG AGGCTATTCA GAAAAACAA AAAAAGCATA TCAAGAAGAT	240
ATACAACATT TTTTTCGTT TTAAAGACT TCAGGTAATG ACAATTATTT GACTGTGGAA	300
CATTTGGATG TTCGGGCTTA TTTGAGCGAA TTATATGATC AGGAATATAG CCGCAATTCG	360
ATTAGTCGAA AAATTGCCAG TTTGCGTTCT TTTTATCAAT TCTTATTAAA AAATGAAGCA	420
ATCCAAGAGA ATCCTTTTTT CTATGTTTAT ATGAAAAAGA AACAGTTACG GTTGCCCCGT	480
TTTTTTTATG AAAAAGAAAT GGATGCATTA TTTGAGAGCG CACAAGGGGA GCAACCGCTT	540
GATTTGAGAA ATCAAGCATT ACTAGAAGTT CTGTATGGGA CTGGTATTCG GGTCAAGTAA	600
TGTGCGAACT TAACGGTGGA CGCAGTTGAC TTTCAAGCTT CTGTTTTATT GGTTCACGGA	660
AAAGGAAATA AAGACCGTTA TGTGCCTTTT GGTTCATTG CACAAGACGC CTAAAGGAT	720
TACTTAGAGA ATGGCCGCGC GCTATTAATG ACAAATATC AAAAAAACA TCCGTATGTC	780
TTTGTTAATC ATCATGGTGA ACAAATCAG CCAACAGGGA TTGAATATGT CTAAACCAA	840
CTAATTAAAA AAAGTTCATT AAATGCGGAA ATTCATCCCC ACATGTTACG CCATACATTT	900



GCGACACATT TATTAAATAA TGGCGCCGAT ATGCGGACAG TTCAAGAATT ATTAGGACAT	960
GCTAATTTAT CAACAACCTCA GATATATGCG CATGTGACAA AAGAAAGTTT ACAAAAAAAT	1020
TATCGAACGT TTCATCCTCG TGCATAATGG AGGTAAACGT CTAGGAGGAA CAATAATGGT	1080
TGAATCACAA TTCTACTCTA CAACAATTTG TGCAGTTGAA AAAGATGGAA AATTTGCAAT	1140
GGCTGGTGAC GGCCAAGTAA CAATGGGCGA ATCTGTCTGT ATGAAAGGTA CGGCCAAAAA	1200
AGTGCGCCGT ATTTACAATG ATGAAGTCGT TGTCGGTTTT GCAGGAAGTG TGGCAGATGC	1260
CTTTACATTA GAAGAAAAGT TTGAAGGAAA ATTAAATGAA TACAACGGAA ACTTAACGAG	1320
AGCTGCGGTT GAGTTArCCC AAGAGTGGCG TACACAACAG TCTATGCAAA AATTAGAAGC	1380
TATGCTGATT GTGATGAATA AGGAAGAAAT GCTCTTAGTT TCAGGAACTG GGAAGTCAT	1440
CACACCAGAC GATGGTATTT TAGCAATTGG tTCAGGCGGT AATTTTGCTT TGTCCGCTGC	1500
AAGAgCCATG AAAAATTTTG GCGATAAAGA AATGCCmGCG AAAGAGATTG CTAAAAATGC	1560
CTTAAATATT GCTGCAGATA TCTGTGTATT TACGAATCAT AATATCATTG TAGAAGAATT	1620
ATAAGCGAGG GAAAATCCAT GAATGAATTA AACAAAACAC CAAAAGAAAT AGTCAAAGAA	1680
CTAGATCAAT ATATTGTTGG ACAGCAAGCT GCCAAAAAAT CAGTGGCGGT AGCTTTACGT	1740
AACCGCTATC GTCGCTTGCA ATTAGAAGAA AATATGCAAC AAGATATTAC GCCTAAAAAC	1800
TTACTAATGA TTGGACCAAC AGGTGTTGGT AAAACTGAGA TTGCTCGTCG TCTCGCAAAA	1860
ATTGTCAATG CGCCTTTTGT AAAAGTTGAA GCAACCAAAT TTACAGAAGT AGGCTATGTT	1920
GGTCGGGATG TCGAATCAAT GGTGCGGGAT TTAGTAGAAA ATGCGATCCA AaTTGTTGAA	1980
AAACAACAAT ACAGTCGCGT GTATGCGCAA cATTAAAAAA GGCCAATCAA CGTTTAGTCa	2040
AAGTATTAGT GCCTGGAATT AAAAAAGAAC AAAAACAGGC TGGCGGTaA CAATTTGAaC	2100
AAaTGATGAA CATGTTTAAC ATGGCTCAGC AACAGCAAGA AGCACAAGAA GAAGTAACGG	2160
AAGATATTCG AACGAATCGC CGAACAATTT TGAACAGCT GGAAAAAGGT CTATTAGATA	2220
ATCGTGAAAT AACAAATGAA ATTGAAGAAC CGAAGAAAAC GATGCCAGCT ATGAACAATG	2280
GCCTAGAACA AATGGGTATC GACTTAAATG AAACGTTAGG CGCTCTGTCA CCAAAGAAAA	2340
AAATCGAACG TACTGTAACG GTGAAAGAAG CACAAGAATT ATTAGTGAAA GAAGAATCAG	2400
CAAAAATTGT CAATGACGCT GATATTCATA GTGAAGCTAT TCGTTTAGCT GAATCAAGCG	2460
GAATTATTTT TATCGATGAG ATTGATAAAA TCACCTCTAA AAGTCAACAA AATTCGGGCG	2520
AAGTCTCTCG TGAAGGAGTA CAAAGAGATA TTTTGCCGAT TGTTGAAGGC TCCCAAGTTA	2580
ACACGAAGTA TGGTCCTTTA CAAACGGATC ACATTTTATT TATCGCTTCA GGTGCTTTCC	2640
ACTTGTCAAA ACCAAGTGAC TTGATTCCAG AATTACAAGG CCGCTTCCCA ATTGAGTTG	2700
AATTAGATGA TTTAACGGCG GATGA tTCGT AAGTATCTTA ACTGAGCCAA ACAATGCTTT	2760
AATTAAACAA TATGTAGCAT TAATTGGCAC AGAAAATGTT TCAGTCATCT TTACAAAAGA	2820
AGCAATTGAA CGGTTAGCAC ACATCGCTTA TGATGTAAAC CGTGATACAG ATAATATTGG	2880

GGCGCGTCGT	TTACACACAA	TTTtagagCG	TTTATTAGAA	GATTtATTAT	ATGAAGCACC	2940
AGATATGCaA	ATGGGTGAAA	TTACGATTAC	CGAAGCATAT	GTCAATGAAA	AATTGAATGA	3000
CATTGTTCAA	AATGAAGATT	TAAGTCGTTA	CATCTTATAA	AAAAGGCTGG	AGGAGAAAAG	3060
ATGGCTACAT	TATTAGAAAA	AACACGTCAA	GTGAACGAGC	TTTtGCAAAA	AAACAATCTG	3120
TTTGACGTAC	AAGCAGAAct	TCCGTACAAC	AAAATGGCAA	TGATTTTAGG	GGATATTTTA	3180
GAAAGTAATG	CGTATATTAT	TAGCAGTTCA	GGGGATTtGT	TGGGGTATAC	TGAAAAATTA	3240
GATGTTAATA	ATGCCCCtAT	CAAAAATATG	TTTAAAGAGA	AAAAATTTCC	TCAAGGATAC	3300
ACCGAGGCCG	TGGATATGCT	CAAAGTAACC	GAAGCCAATA	TTCCAATAGA	TAGTGATTTA	3360
ACTGCATtTC	CTTTTGAATC	ACGAGAGCTG	TATCCGTTTG	GTTTAACCAC	TATTGTGCCT	3420
TTGTACGGTG	CTGGCAAACG	CTTGGGTACG	ATTATTTTAG	CGCGTGTtGA	AAAATCCTTT	3480
AACGAGGATG	ATTTAGTGTt	AGCTGAATAT	AGTGCGACCG	TTGTCGGCAT	GCAAATTTTG	3540
TATCATCAAT	CCAGAACAAT	TGAAGCAGAA	GTTCGCAGTG	CGACTGCCGT	GCAAATGGCC	3600
ATCAACACTC	TTTCTATAG	TGAATTAAAA	GCCGTCCATG	CCATTTTtGA	AGCCTTAGAT	3660
GGCGAAGAAG	GTCGTTTAAC	TGCTTCAAGT	ATTGCAGATG	AAATAGGCAT	TACTCGTTCC	3720
GTAATTGTGA	ATGCCTTAAG	AAAATTAGAA	TCAGCTGGAA	TTATTGAATC	ACGTTCTCTT	3780
GGTATGAAAG	GGACTTATTT	AAAAGTATTA	AATCAACAAT	TTATTAAAGA	GTTGGAAAAA	3840
TAAGCAATTG	CTCGTATCAA	CTAGGAGGAC	AACGCAATGA	CTGTACAAAT	TGAAATGAA	3900
TTTCTAATCG	CAACATTTGC	TGAAGAAGGA	GCAGAATTAG	TAAGTTTACA	ATCAAAAGAA	3960
ACCGGCATTG	AATATATTTG	GCAAGGAAAT	CCTGAATTTT	GGGCCCCGCA	TGCGCCCGTA	4020
CTTTTTCCGA	TTGTCGGTCG	TCTAAAAGAA	GATACGTATA	TGTACCAAAA	TCAAGCGTAT	4080
CATTTAACTC	AACATGGTTT	TGCTCGTGAT	CAAGTATTTG	ATGTGATTGA	AAAAGGTGGA	4140
GAAGAAGTTT	CTTTTtCACT	AAAATCAACA	AAAGAAACCA	AGAAGAAGTA	TCCTTTTGAT	4200
TTTGAAGTAG	TGATTACTTA	TACATTAGAA	CATCAAGAAT	TGACTGTAAA	CTACCAAGTA	4260
GAATACACAG	GGAAAGAAGA	GATGTATTTT	GGGATTGGTG	GTCATCCAGC	CTTTAATGTA	4320
CCTTTAGAAT	CTTCACTAAC	GTTTGAAGAT	TATTATCTAA	GTTTTTCACC	TAAAAAATCA	4380
CGGACACAAA	TTCCCTTAGC	AGGCCCGTTT	ATTGATTTAG	CGAATAAAAC	GCTAGCCCAA	4440
ACGAATACCA	GCTTTGACTT	AACGCATCAA	TTATTTGAAA	ATGATGCAAT	GATTTTTGAA	4500
ACAAAAGGTC	AAACAGCAAT	TACGATTGCC	ACAGACGAAA	GTGACCATCA	AGTCACTTTG	4560
AGTTATCCAG	AAATGCCTTT	TGTGGGGATT	TGGTCACCGA	CACCGAAAGA	AGCGCCTTTT	4620
GTCTGTATCG	AACCTTGGTG	CGGGATTGCT	GATGCTGTCTG	ACGCAACTGG	CCAACtTGCT	4680
GaAAAAtTTG	GCATCaATAA	GTTmCCAGCT	AACGAACTAT	TTAAAmCACA	GTATATGATT	4740
AGTGTAATAAT	AAAGAAAAAG	GCTGTGAACA	TTATTGTtTC	ACAGCCTTTT	TCTTTATTTT	4800
TTTTGCACGT	AAGCCAAAAG	AAAGACGACT	TTCTGTtCCA	TTTTTAATTC	GCTGAATGTT	4860

TTCACGATGA CGGACAAAGA TAAAAGTAGT CAAAGCAATC GCAATCACGG TTAATAACCA	4920
GTTAAATGTT GGTAAAATAG CTGGGACAGT GAAAGGCAAA ATAATAGTAG AAAGTGTAAT	4980
TAGTACAGCG CTAATCATAC TTGTTAAACT CACCATACTG GTTAGATACA GACAAATCAC	5040
AAAAATAAGA GCAGAATAAA TAAAAAATGT GGGGCTGTAT GCTAATAACA TGCCAGCGCT	5100
AGTGGCTACG GCTTTACCAC CTTTGAAATT GGCAAAAATA GGAAGGTAT GCCCTAAAAC	5160
AGCTGCTACC CCAAAGAAGA GCGGATTCAC GCCTTGTAAG CCAAACAAAT AGGGTAATGA	5220
AGTGGCTAAC GTTCCTTTCA AGATATCCAT TAATAATACC GTAATTCGG CAGGTTTCCC	5280
TAAGACACGA AATGTATTGG TTGTTCTGT ATTCCCCTC CCAAATTGGC GTATATCTTT	5340
TTTAAAGAAA AGTTTACCAA TCCAAACACC TGAGGGAATC GAACCTAATA AATAGGCAAC	5400
AAGTAACAAA ATGACGATTT TCATCAAGTT AACCTCTTTC TTATTCAAAC TAGGATTTAT	5460
TTTATCATGA ATAGAAAAAC GGAACAACGT CTAATAACTA AAACAAGAAA AGTTTTTCTT	5520
TTGCTTGAGG AAGCGTTGAC AAATGGTTGA TAACCGTTAT AATAAATGTG TGTCATTAGA	5580
CAAAGACGTC TAATCACTGG TTTGAaAATT AGTTATCAAC GAAGATACAA TATATATGTA	5640
CTTATGAAGT GTCAGGAAA GAGGTGCGAA TCCTCTACAG ACCTACCTAC TGTATGGTGG	5700
ATGAAACCAA TAAGACCACA GATTATTCTG GAAGGATTGG GAGTAAGAAG AAGCTAAGTC	5760
AGGATACCGG CTTGATAAGT CTAATCATTC TTTGTAGGA AAAAGAAGGG GCTTTGAAGA	5820
TAAGTGGCTG AGTATATCAT TAGCATAGTT AAGTGACTGC TTTTTTCTA TTGGAGAAAA	5880
AAGTGGTTTT TTTGTATTGT TTTGACGTTG AGACAAAGGA GGTCATTTC AGAAAATTTT	5940
CCCCAAAATA AAATAGACGA ATGCGAGGAT GAAAAAATGA AAAAATTTAC TTTAACAATG	6000
ATGACTTTAG GTTTAGTAGC AACACTTGGC TTAGCAGGAT GTGGTAAACA GGAAAAGAAA	6060
GCAACTACCT CTTCTGAAAA AACAGAAGTA ACGTTACCAA CCAAAGACCG TAGCGGCAAA	6120
GAAATTACTT TACCCAAAGA AGCAACCAA ATTATTTCCC TAGTGCCATC AACAACAGAA	6180
GTGATTGAAG ACTTAGGTAA AACCGACCAA TTAATCGCAG TTGATACTCA AAGTAGTACA	6240
ATGATGACTG ATTTAAAAA ATTACCACAA ATGGATATGA TGGCTGTCTG TGCCGAAAAA	6300
TTGATTGCCT TGAAACCACA AATTGTTTAT GTGAATGACA TCAATTTAGC TAGCTCAGAA	6360
AGTGTGTTGA AGCAAGTGA AGATGCTGGA ATTACAGTCG TTAATATCCC CACTAGTACA	6420
AGCATCAAAG CAATCAAAGA AGACGTCCAA TTCATCGCTG ATAGCTTATC TGAACATGAA	6480
AAAGGACAAA AGTTAATCAA AACAATGGAT CAAGAAATCG ACGAgTAGCG AAAATTGGTA	6540
AAACTATTAA AAaACCAAAA ACTGTTTTGT TTGAAGTGGC TGCCTTACCA GACATTTATA	6600
GTTTTGGTAA TGGGACATTT TTAAATGAAT GGATTGAACC AATTGGGGCA AAAAATGTTT	6660
TGGCCAATGA AAAGGCTGGT TnCCAGTGAC A	6691

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20072 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

TTAnGGTAnC CAATCGTTCC CCGGTAAATT TGGACAGTTT TAATCTGTTC CACAAACGAT	60
nAACGGCTCC GTTTGTTAAT AACTGCCGTT CTAACATCTT TTCTAAATTG TGAAACAATT	120
TCTTGTTTGG CATAGGGCGG ATCCAATAAG ACCAAATCAA ATTGTAAGTT CTCTTCATAA	180
AACTGTTCTA AGGCACGATT AGCATCCATT TTTCTAACTT CAAATTTTTC TGGTTCTTTT	240
GTAATCGCAA TATTTTCTTT AATTACTTTT AGTGCAGCAA AATTCTTTTC AATACAAATA	300
CTTTTGTTCCA TTCCACGAGA AACGGCTTCA ATCGCTAACC CGCCACTTCC TGAATATAAA	360
TCCAGCGCCA TTCCGCCATC GAAATATGGA CCAATCATAT TAAAAATAGA TTCTTTTACT	420
TTATCTGTTG TGGGCCTAGT ATTGTCGCCA TCTAACGCTT TCAAGCGCCG GCCACCATAT	480
TCTCCAGAAA TAACACGCAT TGTTTCGACT CCTTTATCAA TCATTCTTTT ATAATTCATT	540
ATAACAAAAA GAGTGAAAAA TGTTTGAAAT TTCTCATAAC ATTTCTCACT CTTTTCATT	600
AGACATTCTC TGCTAAGCGA ATCTCTGTAG ATAATTCCTC TACTACATAG CCGTCGTCTT	660
CTTGTTCTTG ATAAGCCGCT TTGGTACCAA TTTTTCAGC AAAATTCATT TCGACATCTG	720
GTCGATACGA TTTATCGACA CTACGAACAA AATGTAAGTT ATTAATTTT TCCATGTTAC	780
TTTCAATATT TTCTTCATTC ATGTAAATCA CCACATATTT CATTTTTCGT GAAACATAGT	840
GAATCAGTCC ATATCGGCGT AACTGTTTTA ACGGTTTTAA TGTGTATACC CAAACAATTA	900
GGCAACGACG CTCTTGAAGC GTAAATCTT TTTCTTCTT AGCTTGCATG GCAGCTTCCT	960
CCACAACCTG AATGTTTCTT TCCTCGTTCA AAGAAAGGAT TTCCAGCATC TACTTTAATA	1020
TCTTCAGAAA TTAGTTGCGC AATCTTTAAA CCAATTGTAT CTAGCAACGT TTGAACGTTT	1080
GTCTCAGCAA AACGATAATT TGCAACGATT TCATTTAAAT CAAGCGCCCG CTCGCTTTA	1140
CGTAGGGCTC GTTGTTCCTC CCGAAAATCG GGTGCATGAA TTCCATAAGC CTCTACTCGT	1200
TCAAAGGCTG ACTTAGCTTC TAAAAATGCT TTTTGTAAGT GAGCAACTTC CGCCGAAGCA	1260
TACATAGCTT TTTTGTGTTT TTTATAGTTT TGAACAGCAT CACTTTCTAA AAGTGCTGAA	1320
ATTAAACATT CTGTTTGATC TTCAATTGCA AATAAATGTT CTGTTACAAT CATAACTTTT	1380
CACTACTCTC CGTGATTGTC CTATTTTCTG GTCGTTTCTC TAATAAGACT AACACACCTT	1440
CATTGGAAAA GGCAACGCCC AATTCTGTCT TTTCTGACTC AGCAAAATAG CTATCATATT	1500
GCGGGTCCGA AAACCATGac AAAATAGTAA AGGTACTATC ATAAACAGGT TCTTGAAAA	1560
TACCTGcTGG TTCTCTCAAC GAAAGTTkGC TCGTTTTAAC TAATAAAGrA AATTCGkCCT	1620
TAGTCmAATG AAaGGGATGA ATGGCTGTAT GTTCTCTCTC CGTTTGTAAC AATTCTGCTT	1680
GTCGATTGTC CGACAAAATT GTTGGTGTAT GGTTAGCCAT CGTTTGATAC AATGTTTGGG	1740

cATCTTTTTG GCTTTGTGTA GAAACTGAAA AAGCAACTTG ACTTTCTTGc TGTTTAACCA	1800
GATTCATGAC ATAAATCGCT TGGTTACTTT GAACAGGATC GAAGGTCATT GACTGCGTTA	1860
TATCAACTGG CAAAGGAGAA CCTTCATTTA ATTGATACGG CATTTCCGTG AGCAACGTTT	1920
CTTTATTGAC ATAATTAACCT GCCATTAAAC CGCCTGTCCG CTGATTAAAA AATAACACCG	1980
CAAACTGCC ATTATCAAAA GCAACCAGTG GTCGATAGTT CATATCTTCT TCCATCAGCT	2040
CGACATCGTA AGCTTCATGA TGAAAATTAA ATGCAAAATT CGAATAAATC GTCATTAAAT	2100
CAGACAAGTC AGCAAGCTTC ATATCTATTT TAAACGGTTC TATCATTTTG TCATTGTTAA	2160
AGGCCTTAAT TGCGATGACT TTATTTTCTC GAACGTTGAC TTCTAAGTAA TCATTTCATCG	2220
TCAAACCGAA CGTCCATAAT TCATACGTAC GACCTGTTTT AATTTTTTCT TTAGGATTGC	2280
CAAACGGCG GATAAAGATT TCCACATCTT TCCCTATGTA TGCTGCGTAC CAGAAGCTGG	2340
CAATTCTTCA TGAGGAACAG CGGTGTGAGC GCTGGATAAG CGCCCACTTT GAGCTGTTTC	2400
GGTTGGTTTT TCAGCAGGGA AAAAACTGG CTCTAAATAT CCAATGAGTA GAACTGTTAA	2460
AAAAACAGCA GTAAAGGCC AAAATCGCTT CATTTTTTCC CCTCTCATCT TATTTTTCTT	2520
TCACTTCTAA TAACAAATCA CCTGAGCTGA TTGcTTCGCC TTCTTCAACA TAAATGTGAT	2580
CTACCGTGCC CGCAAAACGT GCTTCGATAG TCGTTTCCAT TTTCATGGCT TCTGTAATCA	2640
GCAACGGCTG ACCTTTTTTCT ACTTTATCGC CACGTTTGAC CAATACTTGT AGAACAGAAC	2700
CAGACATCGT TGCGCCGATT TGTCTTTTAT TAGTTGGTTC CGCCTTTTGT TTTACTTGAA	2760
CAGCAGACTT AATAGAGGCA TCTTTAACTA AAACCTCAGC ACGTTGCCCA TTCAAATTGA	2820
AGAACAAGAC ACGGT'TTCCG TCAATATCAG GTTCGCCAAT TTCATCTAAA CGAATAATTA	2880
ACGTTTTTCC TCGTTCAATT TGAACCTCCA ATGTCTCTCC TTGACGAATA CCATTAAAGA	2940
ATGTTGGCGT ATCTAATAAG GTTATATCTC CAAAGGTTTC GTATTTTTGA CGATACTCTA	3000
AAAATACTTG TGGATACATT AAATAACTTa AAACTTCTTC CAATTTAGGT TGaTACCCAA	3060
TTTTTTCGGC TAATTCTTCT TGTACTTTCG CAAAATCAAC AGGTGCTGcT AAGTCACCTG	3120
GcCGCTCTGT GAAGGCTGGT CGACCTTTTA GAATAATTCTG TTGCAGCTCT TTCGGGAACC	3180
CACCAACTGG TTGACCTAAA TCACCTTGGA AAAAAGTCAC AACTGATTCA GGAAACTTA	3240
GTTCTTCACC ACGCGCATAA ACATCTTGTT CTGTCAGATT ATTTTGAACC ATGAAGAGTG	3300
CCATGTCTCC CACGACTTTT GAAGATGGCG TCACTTTTAC AATATCGCCA AACATCAAAT	3360
TCACTGTGTG ATATATTTTT TTGATTTTCT CCCAACGGTG CCCTAAGCCT ACCGCTTTTG	3420
CTTGCTGCTG TAGATTAGAG TATTGCCAC CAGGCATTTT ATGCATATAG ACTTCTGTTT	3480
GCGGGGCATT TAACCCATTT TCAAATGGTT GATAATACAT GCGCACATCT TCCCAATAAT	3540
GATTGATTTT CTGTGCATTA TCAATGTTAA TAGTTGGCGT CCGTTCACCA TTGACTAAAG	3600
CATAATATAG ACTGTTTATA CTTGGCTGGC TAGTTGCGCC ACTCATCGCA CTCATTGCGA	3660
CGTCAACAAT ATCAACGCCC GCTTTAGTAG CTGCTGAATA TGTGATGATC CCATTGCCAC	3720

TAGTGTCTGTG AGTGTGGAGA TGAATTGGTA AATCCGTGGC TGCCTTTAAT TCACTAATTA	3780
AACGATAAGC AGCTTGTGGT TTCAATAAGC CAGCCATATC TTTAATCGCA ATGATTTGTG	3840
CACCTAAATT TTCCAATTCT TTAGCCATAT CAAGGTAATA TTGAACATTA TATTTTGCTC	3900
GGGCTGGATC ATTGATGTCC CCAGTATAAC AAATTGCTGC TTCCGCAATT TTTCCGGTAT	3960
CCCGAACGAC TTGAATACTT TTTTCCATTT GAGGAATCCA GTTTAAGCTA TCAAAAATTC	4020
GGAAACATC GACTCCTTGG CGAGCGGATT CTTTAATAAA TTCTTCAATA ACATTATCAG	4080
GATAATTTTG ATAGCCTACC GCATTAGATC CTCTGAACAG CATTTGTAAA AGTGTGTTTG	4140
GCATCAACTG ACGAATTTTT CTAAACGTT GCCATGGGTC TTCGGTTAAG AAGCGATAGG	4200
CAACATCGAA CGTAGCGCCA CCCACATTT CACTAGAAAA CAGCTCAGGC AGGGCTGCAT	4260
CAGTTAGACC AGCAATTGCT TTAATCTT GTGTTCTCAC ACGAGTAGCT AATAAACTTT	4320
GGTGCGCATC ACGGAACGTG GTATCTGTCA TTAACACACT TTCTTGATTC TTAACCCAAT	4380
CGATGACGGC GGTGCTCCT TGAGCGTCTA GAATATTTTT AGCTGTAATC ACTTTTTCTG	4440
GAAC TTCAAT ATCTGTCGGC ACGCGCGGGG CTTCAAATA TTTTTCTCC GTTCTTTCAA	4500
TGCCAGGGAA CCCATTGACC GTTACTTCTC CAATGTATTT CATCGTTTTA TTGCCACGAT	4560
CGCGCATACG AGGAAATTCA AATAATTCAG GTGTATTATC AATAAAGGTT GTTTTGGCTT	4620
CGCCAGATTG AAACGCTGGA TAGCTCACAA CATTTTGTA A AACGGAATA TTTGTTTCA	4680
CGCCACGAAT CCGAAATTCT TTTAAGCAGC GTTGCATTTT ACTGATTGCC TGTTCAAAG	4740
AGAAACCATG CGTACAGACT TTAACCAATA AAGAATCAAA GTAGGGAGTC ACAGCGTAGC	4800
CAGAATAAGC ATTCCCTACA TCTAAACGCA CACCAAAACC ACCTGGTGAA CGATACGTAT	4860
CGATTTTCCC TGTGTCTGGC ATAACTGGT TCAGGGGATC TTCAGTCGTA ATCCGACATT	4920
GAATAGCCGC GCCTTTTAAT GTCAATTCGT TTTGTTTGG CAAATGCATA TCTTTATGCA	4980
AATCAAGCCC TTGCGCAATT TGTAATTGAG AGATCACAAT ATCAATATCT GTAATCATTT	5040
CTGTGATAGT ATGTTCTACC TGAACACGAG GATTCACTTC AATAAAATAA AACTGATCCC	5100
CTTCTACTAA AAAC TCAACC GTTCCCGCAT TCACGTAGCC GACATGTGCC ATTAAC TGCA	5160
CAGCAGCCGA ACAAATGGCT GCTCGTTGTT CTTCATTCAT TGATACACAT GGTGCGACTT	5220
CTACCACTTT TTGGTGGCGC CGTTGAACGG AACAAATCAG TTCAAATAAA TGCAAGACGT	5280
TCCCATGATG ATCGCCTAAA ATTTGTACTT CGATATGTTT AGGATTAGAA ATATACTTTT	5340
CAACATAAAC CTCGTCAGAA CCAAAGGCTG CTTTCGCTTC ACTTTTTGCT CTTTCGTAAC	5400
CTTCTCGTGC TTCTTTGGCA TCGTGGGCAA CGCGCATCCC GCGACCGCCG CCACCTAAAG	5460
CAGCTTTAAT CATGATAGGA AAGCCATGTG TTTCACCAA AGCTACAACC TCTCCACCG	5520
TTGCTACTGG CCCGTCTGAG CCTGGAATCG AAGCAATCCC CGCAGCTACA GCTGCTTCTT	5580
TCGCTTTAAT TTTATCGCCA AAAATATCTA AATGATGTGT TTTAGGTCCG AAAAAATGA	5640
TTCTTCTTC TTCACAACGT TCGGCGAAcG TAAGTTCTCT GATAAAAAGC CATAACCAGG	5700

ATGAATGGCA TCIGCTCCCG ATTTTTGGC AATTTGGATA AATTTTCAA TGTCTAAATA	5760
GGCTTCAATT GGTTTTTTCC CTTTACCAAC TAAATACGCT TCATCTGCTT TGAAACGATG	5820
AACAGAATAC TCATCTCTG CAGCATAAAT CACTACTGTA CCGATATCTA ATTCTGTACA	5880
AGCTCTAATA ATTGGAATCG CAATTTGGCC AGGATTGGCA ACTAATACCT TTTTCATAAC	5940
GTCACCTCTA CCATCAATTT CTTTCAAAT TATTGGTTCT TTCTTAAGA TAAATAGTAC	6000
TCATTTTCTA ATTTCTGAGC TGTTCATCT GCACTGATAT TCAGGACAAA GGCCACAGCA	6060
ATCGAAATAA TCAGCAACT ATTCCCACT TCACTCAAGA ACGGGAAAGT AATTCCTGTT	6120
AATGGAATAA TCCAGIAT CCCACCACT TTAATAAGA CTGAATTAA TAACATTGTA	6180
CCAATTCCTA TACACATAG TAAATTAAA GTTTCTTCTG ACGTACACC GACTAAAATA	6240
ATTCGCGCAA TCATAACAT CAACAATCT AATTCGCTA GACCGCCGAT AATGCCTAAC	6300
TCTTCTAAG TATCGCAA GATAAGTCC GTATGTGCTT CTGGTAAAA GCCTTTTTTC	6360
TGAACACTAT TCCCTAGCC TTGCCCAAC CAGCCGCCAT TACTGATTGC ATAATAAGAA	6420
TTAGCTAACT GATGCTTAA GTCCGTTCA TCTAAAAAG GATTTTAAA AACAGCAAAA	6480
CGATTATAAA TATATTGTA ACGCGCTGGG AATTTTTC CTTAGACAT AATTAATAAC	6540
TGAATAGCCG TGATACIAC TAAATGCCC AGCCACCGA CCAAGTAGGT ATACATATAA	6600
TTAATCCAC TTGCTAAAC CATCACATG GTAATTAAG TTAGAATGGC CGCATTCCCG	6660
AAGTCAGGCT GATTTGCAAC TAGAGCAAT AATACAAAA CAGCATCAA CGGGCGCCCG	6720
GCAGCTTGTT TAAATTGAT CATGCCCTCA TTAATTGCT TTTGTGCCG TGCTAAAATA	6780
TAAGACAGAT ACCAGACAC CATATTTCT AATACTCAG CTGGTTGCAT CGAAATCCA	6840
CCAATCTCAA TCCAACCGC TCGCCCTTA ATTTCTTCC CATACCCAGG AATCCGAACG	6900
GCTAAACCA TACCGTAT TACCGCATG GCAAACATG TAAACTTCG ATTCTGAAAA	6960
ACGGATGTTT TCATCTATA AATAAATAC ATTGCAACTA ACCCTACTAC CAAAAGGCA	7020
ACCTGGTTAA TACGATGA GGTAGGTGA AACCTTTCA TTAATTGTA TCGGATGTG	7080
GAACTATATA CCATAATCA TCCACGATA CTAAATTA AATATGGAAT AAAAATACTA	7140
TAGCTAGCA AATGCCCTT CTTTACTTG TTTGGCAAG CTTAACCTC CTTGCCAATC	7200
TCGTTTTTTT CTGTGTTTC TCATAACA TCTGTATATA ATTGATTCAA TTCGCTTCT	7260
AAATCACTTA GTAAGCATG TCCCTCTCA TTAGTTAAA TCCCCAACG GGTGCGATAG	7320
GTCACCTGTC TTGAATACC AATCATTTG GTATCCACCA CTCTTCTAA TGCTTTACAT	7380
TGAGAAATGC ACAAATCT TTTTGTATA CCAATAAGCA TTTAATGCG TTCTGCATCG	7440
TCTTCTAATA ATTGAATGGC AATTTCTTT GAATTCCTT CTTCCATTG TACAGAACCT	7500
CCCTATCTTA ATTAATAT TATACATA TCTCAGTA GATACATGAT GATTCAGAA	7560
AATTTCTTT AAGAAATTT CTGAACCAA GCGATATAT AAGTAAGAAA TCTCTTTATA	7620
TAACAAGCTT ATTTATCAC TACACACA AATACATTA AAGACTTTTT TCTTATTTTC	7680

TGCTTTTAAT TTTAAATAAA ATAACAACT ATATCATTTT TTTGTCTCTT GATATTCTTT	7740
AATTACTTCT AAAAAGGCGG TGCCATATyT ATCTAATTTA TTTTGGCCGA CCCCTTTTAT	7800
TkGCAATAAT TGAATCGTTG TTTGCGGTAA TTTTTCACAC ATTTCTTTAA GTGTACTATC	7860
GGAAAAGACT ACATATGGCG GCACGCCTGC CTCTTGAGCC ATGTCCATAC GTAATTCGCG	7920
TAATCGTTCA AACAAGGCAT CATCTACTGC TACTTTTCGA ACTTTTTGGT CTTCTTTTCG	7980
GAAGACTTTC CGTTCCCCTA AAAGCACCTG AACACCTTCT GTCGAGACTG ATAACAACGG	8040
AAATTGCCCA TCAGAAGGAA TTAAATAACG TTCTGCTGTC AAATAATCAA TCAGTTGTGT	8100
GACTTCTTTT TCGGTGCGAC CTTTCATTAG TCCGTATGTT GGCAAACGAT CAAAATGCCA	8160
TTGATCAATT TTTTGATCTT TTGAGCCAGT CAAGACTTTA CCAACTAAGC CCTTACCGAA	8220
ACGTTCTCCC ATACGTTTTA CACAAGACAA AACTTTTTGG GCATCGACCG TAATATCGAC	8280
TAATTCTCGA TTATCCAGAC AATTGAGCA GCGACCGCAG TCCGTCCCTT CTTACCAAA	8340
ATAGCGTAAA ATATATTTTT GTAAACACAT TTGTGCATTC GCATATTGTG ACATTTCCCG	8400
TAATTTCAAG TACTCTTTCT GCTTATAATC AATGGTCATT TCAGATTGTT CAATAAAGTA	8460
TTGTTGAATT TGCAAGTCTT GTGGAGCAAA CATTAAACACA GCATCACTAG GCAGACCGTC	8520
ACGACCAGCT CGGCCGGCTT CTTGGTAATA AGATTCAATA TTCCCTGGTA CTTGTGCGTG	8580
AATGACAAAA CGAACATTAC TCTTATTAAT CCCCATACCA AAGGCATTTG TTGCTACCAT	8640
CACTTGCACT TGATCATACA AGAATGCTTC TTGATTTTCA CTACGTAAC GTTCACTCAT	8700
GCCCCATGA TACATCCCAG CTGCAATTTT TTTACTTTCT AATAAATGAT AAATTCGTTT	8760
AACTTCCTTC CGCGTGCTTG CATAAATAAT ACCTGATTGT CCAGTATTCA TCTTTAAATA	8820
TTCCAAAAGA AAAACGTCTC TATTTTGATC TTTCACGACT TGAAAAGATA AATTTTCTCG	8880
AGCGAATCCT GTTTTGATTT CATTTTCTGA TGGAATTCTT AATTGCTTGA CAATATCCTC	8940
CGCTACTTGT GCGGTTGCTG TTGCTGTTAG TGCAATCACA GTTGGTTGCT GTTGAAATTG	9000
ATCGATAATC TCTGCTAAAC GCAATAGCT TGGACGAAAA TCATGCCCCC ATTGTGAAAT	9060
ACAGTGTGCC TCATCTACTG CCAATAAATC AATCGGTACA TGTGTCAGCA TCTGTTGAAA	9120
ATCATAAGAT TCTAATCGCT CTGGCGCAAC ATACAATAGC TTGACTTCTT TGTTAACCGC	9180
TAATTGAATA CGATGATTCA TCTCTTGATA GGAAATGGTA CTATTGATAT ATGTAGCTGG	9240
AATCCCCATT AAATTCAGGG CGTCTACTTG ATCTTTCATT AAAGAAATTA ATGGTGAAAT	9300
CACCAATGTT AAATTATCCA GAAGTAAGGC TGGCAACTGA TAACAAATCG ATTTGCCTCC	9360
ACCTGTTGGC ATAATGCCTA AGACATTTTC TTGTCGCAAT ACATGACGAA TAATCGTTTC	9420
TTGTCCTGGC CGAAAATCAT CATAGCCAAA TGTATCTTTT AATAATTCTT GTAATGCACT	9480
CATTTATTGC GCTCCAATCT TTGTTCTACA AGTCATTTTA ACACGCTAAT CTAACAGGAA	9540
CAAGGTGTGT TAAACAAGA CACAATCGTG AAACACCAGT TAACGTGAAA CATCAAAAAG	9600
AATGAAAAAT TGTGAAAGCT TGCTATCAAA AGCTTTCAAA AATAGCGGCA TTTGCTTGTT	9660



GTGGATAACG AGTAGTTATC CACACATAAA AAAATAACCT TATCTTAAAA TACGCATTAA	9720
TTCGTTTTTG CCATCAAAAA AACCAACTAT TTTCTAGTA AAACGAGCAT AGTTGGTTTA	9780
AATGTTTATT TGTGAAAAA TTATTCAGCT TTTTCTTTT TCTTGCTTGC TTTTTCGCGT	9840
TCGTTTTTGT TAAGAATTG TTTTCTTAAA CGGATGCTTT CTGGTGTTAC TTCACAGTAT	9900
TCATCTTCAT TCAAGAATTC TAATGATTCT TCTAACGTTA ATTTTTTCGG TTTTAAATG	9960
ACTGATGTTT GGTCTTTCGT TGCCGAACGA ACGTTAGTCA TTTGTTTAGC TTTTGTGATG	10020
TTCACAGTTA AATCGTTGTC GCGGTTGTTT TCCCAACAA TCATTCCTTC GTAAACTTCT	10080
GTCGTTGGTT CAACGAAAAC TGTACCACGT TCTTCAATAC TCATGATTGA GTAAGTCGTT	10140
GCTTTACCAG TATCGATTGA AACTAGTGCA CCTTGATGAC GGCCACCAAT CGTTCCTTGA	10200
ATCATTGGTA AGTATTGATC GAAGGTGTGG TTCATAATAC CATAACCACG AGTCATTGAT	10260
AAGAATTCTG TTGAATAACC AATTAAGCCA CGAGCTGGTG CTAAGAAGAT TAAACGCATT	10320
TGTCCGTTAC CGGCATTAAT CATATCTTGC ATTTCGCCTT TACGTAAACT TAGTGATTCA	10380
ATCACGCTAC CCATGTATTC TTCAGGTGTG TCAATTTGAA CACGTTCAAA TGGTTCACAT	10440
TTAACTCCAT CAATTTACAG TTCAATAACT TCTGGACGAG AAATTGTAA TTCATAGCCT	10500
TCACGACGCA TGTTTTCAAT TAAAATTGAT AAATGCAATT CGCCACGACC TGATACAGTC	10560
CAAGAATCTG GGCCAATTGG ATCAACACGT AAAGATACGT CTGTTTGTA CTCAGCCATT	10620
AAACGTTCTT CGATTTTACG AGCGGTGATG TATTTTCCTT CACGTCCCGC AAATGGAGAA	10680
TTGTAACTA AGAAAGTCAT TTGTAAGGTT GGCTCATCAA TGTGTAGAAT TGGTAATGCT	10740
TCTTGATTGT GAACATCTAC AACTGTTTCC CCAACGAAAA TGTCTTCCAT TCCAGAAACG	10800
GCAATTAAAT CGCCGCTTT TGCTTCATCA ATTTCCACAC GTTGTAAGCC AAAGAAACCT	10860
AAAATTTTCG TTACACGGAA ATTTTTCACG CTGCCATCTA ATTTTCATCAA CGCAACTTGG	10920
TCGCCGACTT TCATTGTGCC ACGGAACACA CGGCCAATCC CAATACGTCC AACGTAATCG	10980
TTGTAGTCTA GTAATGAGAC TTGGAATTGA AGTGGTTCGT CTGAATTGTC AACTGGAGCT	11040
GGCACATGTT CAATAATTTT ATCAAAAATT GGGGCCATTG TTGGCTCTTG ATCTGCTGGA	11100
TCATCTGATT CACTTGAAGT TCCGTTTAAA GCAGAAGCAT AAACAACTGG GAAATCTAAT	11160
TGATCGTCGT CTGCACCTAA TTCGATGAAT AACTCTAAAA CTTCATCTAC TACGTGTTCA	11220
GGACGAGCAG AAGGTTTGTC AATTTTGTTA ACAACCACGA TTGGTGTTAC TTTTGTCTT	11280
AATGCTTTTT TCAATACGAA ACGTGTGTTA GGCATTGTTT CTTCATACGC ATCGACAAC	11340
AAAACAACAC CGTCTACCAT TTTTCATGATA CGTTCTACTT CACCACCGAA GTCCGCGTGT	11400
CCTGGTGTAT CTAGAATGTT GATACGTGTA CCGTTATAAT CTACGGCTGT ATTTTTGCT	11460
AAGATAGTAA TTCCACGTTT ACTTTCAAGT GCATTGGAAT CCATTGCACG TTCTTGTAAT	11520
TGTGTGTGTC CATCTAAAGT GTCAGATTGT TTTAAAAGTT CATCTACTAA GGTTGTTTTA	11580
CCATGGTCAA CGTGGGCGAT AATTGCCACG TTACGAATAT CATCTCTGTA TTTCAATTTA	11640

ATTGCTCCTT ATCTATATAC GCTTATGTTA CTTAAGTCCA TAAAmGCCGA CTGCTTATTA	11700
TmGCAAACTT TTTTCATTAA AAAAAGcTTT CGCCTGCAAA AATAATGAAA GTTTTAGAAA	11760
TTATTCGGCT TTTTCTGAAA ACAAGGACAT TTTTTtACAA ATTTtyCTCT TTTAAAAGAG	11820
GAAGAGCCTT ACCTTTCAGC AACGCTCTTC TTTGATTATT CCGATTCAAT TAAATATCGG	11880
GTAAATACTT CATCATAGGT TTCAGGAGTT GCTGCAATAA AGTATTCACG ACCCGCAAAA	11940
GTTAATGGTT TTCCTGTAAT ACCAGAGTAT TTAAACCCAA ATTCTTCCAA AAGTACTAAG	12000
CCTGCTGCAT AATCCCAAGG ACTTAGATTA GATAAATAGC CATGATGGTT GCCTTTTAAC	12060
ATAGCAATAA TTTCCAAACC TGCACAACCG CTAATACGCA CGCCCATGCT TGCTTGTCGG	12120
ATTTCCGTTG CGTGATGAAT ATTTTTTCCA TGCATGTAAG CATTCAATGCC CCATAATCCA	12180
TCTGCCAAGG CTTTCATAGT AGGCGCTTTT AATAATTGAT TATTCCGATA AACGCCAAGC	12240
CCTTTGCCGC CCCAATAAAG TTCTTCTCGC ATCACATCAT AGATAAAACC TAATTTACCA	12300
ATGCCGTCTT CATAACAGC TAACATAATA CAAAATTTT CTCTTTCCAT CACGAAATTC	12360
ATTGTGCCAT CAATCGGATC GATGATCCAC ACGCGACCAG CAAAACTTT TAAAGTATTG	12420
TAGCCTTTTT CTTCCTTAA AATTTGATCT TCGGGAAAAT TTGTTTGAAT TTTATTCATT	12480
AAAAATTCTT GGGTTTGTTT ATCCATATTT GTCATAAAT CGGTGCGGCC ATTTTTCTGT	12540
TGAACGGTTA ACTCTGATTC CAAGTTCATT CGAATCACAT CGGCTGCTTC AAAAAGCCAA	12600
CTTTGAATTG TTTGAATAAA TTTTTCGTT TCAGTCATCT AATCATCCCC CAGCTTTAAT	12660
TTACCGCTTC CCTGTTTTTT AGCTTGTTGA AGCGTACGAT ATAACGAATA TCCTGAAGCA	12720
TTTTCAAATT CACGACCTAA GCGTTTTTCT TCGCCGATAC TTTTAACGAC CGTTTTAAAT	12780
TGTTGATAAG TTGTTAAAAA TTTGTCAACA GGCAAGCCTG TTTCGTTGGC TTGCTCAACT	12840
GCTGTCCACA TATTAGkGAC AATCACCATT TCCtCTGTG TCCAATCTAA ATCTAATGGA	12900
TATTGATAGT CTTTCATGTC TTCACCTCAT CTCTCTAGT TTAGCATATT TTCATAAAAn	12960
AACATCTTAT TTACGTCGAT TTTTAAATGC ATTAAAGAGA GGATATTCTG CTGTTCCTTT	13020
GCGATACAAA CtTTGTTGTG AAGCATAAAT ACCGCTATTT CCTGAACACA TAAACTGAT	13080
TAAGCAAACC ATGAAAAAAT AGACCGCCGC TTCACTACCA AATAGTTCAA TTCCCATAAT	13140
GAAACAAGCA ATCGGCGTAT TCGTGGCGCC AGAAAAACA CCAATAAAATC CTAAACCTGC	13200
TAAAAAAGGA ATTGATAGAT GTAACAAAGG GGCCAAAGAA CTCCCTAACG TAGCCCCAAT	13260
TTCAAACAAT GCGGTCACTT CTCCTCCTTG GAATCCGGCG CCCAAGGATA ACACCGTAAA	13320
GAATAATTTT CCAACAAAAT CAAAAGGCTG CGCATTACCT GCAAAAGCAT CTTCTAGTAA	13380
AGGTAAACTC AAGCCTAAAT AACGTTGGTT GTTTAAGACG AGCACAAACA GAACGACAAT	13440
GGCAGCGCCA AGAAACGCAC GAAGGTAGAC ATTAGCAAAC CAATTTGCAT AGGTCTTTTT	13500
TAAAAAGACA ATGGAACGAC TAAAAACCCA ACCTGCCAAA CCAAAACAAA TCGATGCCAA	13560
AAATAATTTT ACAAATAGTT CAACCGACCA AGTCGGAATT TTGCCCATCG GATAATGCGT	13620

ATGTGTTACA CCAAACTTT CAGTTACAAA APTAGCAAAT AAGCCTGCAA AAAAATTGG	13680
GAAAAGTGCC TCTGCACGAA CTTTCCAAAC CGCCAAAAC TCTAATCCGA AAACCGTTCC	13740
CGCTAATGGT GTGCCAAAA CAGAACTAAA ACCTGCACTA ATTCCGCTAA TAATAATTAC	13800
TTGTCGTTCC AAAGCGGACA GTTTAAAGAC TTTTCCTACA GCATTTGCGA TTGCGCCCC	13860
CATTTGCACC GCCGTCCCTT CACGACCAAC AGAACCACCA AATAAGTGC G TAGTAATCGT	13920
GCCAAACAAT GTTAAAGGAA TTAAGCGCAA GGAATATCT TCTTCCCCG CATTGCCTTG	13980
TTCAATCACT AAGTTATTT CTCGGCTAGC ATTTTCCCA AAACGCGTAT AGAAATACGT	14040
AAAAACGACA CCGCTAATGG GGAGGATAAA GAGTAACCAA GGATTTGCTA AACGAACATC	14100
TGTTACATAG GTGAGACTCT TTAAGAAAA GCGGATAAC GCCCCATAA ATAAACCAAT	14160
AATTGCGGTA ATGAGCAACC ATTTTGCCAT ATAACGCGG ATTGATAGTA GTTGCTTTGT	14220
TTCTTTTTC ACTTTACATG CTCCTTTACA ATTTTCTTAT CGCCAATCCT AGCGTACAGA	14280
AAAAGCCTAC CAAAGGTAC ACTTCTCCCG TTTGkaGGCG TCATTAATTT ATTCGTCCma	14340
TAAATAGTTC GTGGTGAACA CCATCACCAT GTTGCTATTA AACTAATwAa ATwATACCaC	14400
GCAaCTCTGG AAAGTGCAAT AAAAAaTCaA TTGCGTTCG TTTTtATCCA CGTGCTATTA	14460
CCTTTGATAA ACCGATAATA AGCCATCACC AACTAACGA ATAAAATTAA GTACATGTAG	14520
GAAATAGTAA AAGGCAAAAT AATCAGCCCC AGTTTTGCAG GTACTTTTAG TTGAAAATTT	14580
TTCAACTCGC GATAATAGTT GATTGTAAAA AAAATGCCGA AACTAACGA AATACTGACG	14640
GCCAAACAA AACTAATGAC AAGCGATAAC GAAAAGCCTA ACCTATACAA GTGTTGCAGT	14700
TGAAGCCCAG TTAAGACAAA AACGAGCATC CCCGCTACGA GATTAATAAA GGGTTGCGCT	14760
AAAAAATAAT AGATTCCACA CTTCTGCCTT AGCGTTAATG ACTTTGATTT AATCACCTG	14820
GGTAAATACG CTAGACAATC TAAATTTCTT TGAACCAAC GACTTCGTTG ACGAATAAAA	14880
CGTTTCACAT CTCTCAGTGC TTCTTGAGCC ATAAAGGCCT CGTCAATATA AGCAATCGAT	14940
AACTCTTTCA GCATTAATTT AATGGTTAAC TCATAGTCAT CCAACAAAGC ATTTCCCAA	15000
GGCTGATAGC CTAAGTGTT GGTAAATCATT GATAACCGAA AAACTGACC ATTGCCACCA	15060
AGTCCGATGG TTTTGTGTC TCGACGCGCT TTTGAATAA AATTATTAAT CGTAAAAAC	15120
TCAACATCTT GTGCAATAAA CATAAAATTT GTGACATCTT TCATTTTGAT CCTTGTTTGC	15180
GCAsCACAGA CTGTTGTATC TGTGAAAACA GCAGTTAGCT CATTGACAAT ATTTTCACTA	15240
AAAAAACCAT CTGCATCTAC CACACCCACT AAACTTGAT CCAAGAATA GTGTCCTTTT	15300
TCTGCATCTG TTAATACAAT TTGGAAAGCG TGGTTCAAAG CGGCCCTTT GCCTTGTTGC	15360
GCATTGGGAA ATTGACGAGA CAGAACTTGG ATTGGACCAG CAATACTGTC CATCATTGCT	15420
TTGGTTTGGT CCGTTGATCC ATCATCAATC GCAATTATTT TTTTCTTAAT TGGTAAAGCA	15480
ACCAGTTTTT TTAATGTTTC TCCGATGACT TTCTCTCGT TTACACACGG TACAAGAAAG	15540
TAGCAGAAAA ATTCTTCTGC TACTTCGACT TCTCTTCTTG TTTTTTCTT GTTCGAAAAC	15600

TGAGAAATAA	ACTTGTAAT	AAATAGGTTA	AGAAGATGAA	CATAATTACT	TTGTATAACT	15660
GATACAAATA	AATCATGTAA	TTTGACCTCT	TTTAAAAAAA	TAACGGTATA	GAATGTAGGC	15720
AGTGACAATC	GAAACTAGAA	TAACGCCAAT	TAAATCAACC	ACAAAAAAGG	CGCCTGTTGA	15780
AATTACCCCA	AGTTTCGGTA	AAGCTGATtG	GCTCATtAAA	TACCCCTTTC	TCTCTATTTG	15840
TTTTTACCAA	CACACCACCA	CGATGACCAT	CACACACCCC	CTATAATAAT	ACTTTGGaGG	15900
ATTATAGTAC	ACAGGTGGAT	TATAGTACGT	AAmTGGcGGA	ATATAyCCTG	GcCAATATGT	15960
TGGTGGTGGC	GGATTATAAA	CCGGTAwTTT	TCGGTTTTCT	CGAACAATTG	cAATTGTTTG	16020
ACCCGCATGA	TTTTGCGTCA	GACTGACGCG	TGTTGGCGTA	GTATTCGTG	AATAACCAGT	16080
CGGCGCTTGA	ACTTCTCTAA	TTTCATATAC	CCGGTTGACA	GATAGCCAC	TAATAGTGAC	16140
TGTACCATCC	CAACCTGTTG	TTACTCTTTG	TGTATAGCCA	TTAGTTGTAT	CTTCCATTC	16200
AAAAACAGCG	CCCGCTAAAC	GTGCTTTTGT	TTCAACATCT	TGTTTAATAA	TTCTGATAGC	16260
ACCATAGTTT	GCTTTATTGT	TTAACTACTAA	CGTATACGTT	TTTGTGCCTT	GACCTCGTAA	16320
ATCAATCCGA	TGGACCCTAC	TATCCAACGC	ATACCCAGTT	GGCGCTTGTA	TCTCGGTGAT	16380
TTCATAAATA	CGATCCGACG	ATAATCCTGG	AATCAGTACT	TCACCATTTC	CATTGCTGT	16440
AAATGTTGCC	GTATAGTTAT	TGGTAGTATC	GCGAAATCTG	AATTCGCTC	CTGCTAAACG	16500
TTTTCTTGTA	TCCGTATCTT	GCTTAATAAT	TTTGATGGCG	CCTTTTTGTG	CTTGATTATT	16560
TAAGACTAAT	GTATAGGTTT	TACCACCTTG	TCCACGTAAA	TCTACTTTAT	GAACACGTCT	16620
ATCTATGATA	TAACCATTCTG	GCGCTTTGAT	TTCTGTGATC	CCATAGATAC	GATCACCGGC	16680
TAATCCAGTA	ATCAGTACTT	CACCGTTCCT	ATTCGCTGTA	AACGTTGCAG	TATAATTATT	16740
GGTTGTATCT	CGGAATTTAA	ATTGTGCGCC	TGCTAAACGT	TGTTTCGTAT	ATCCATCTTG	16800
CTTAACAATT	TTGATGGCAC	CTTTTTGTGC	TTGATTATTT	AAGACTAATG	TATAGGTTTT	16860
ACCACCTTGT	CCACGTAAAT	CTACTTTATG	AACACGTCTA	TCTATGATAT	AGCCATTCGG	16920
TGCTTTTATT	TCTGTGACCT	CATAGATACG	ATCACCGGCT	AATCCAGTAA	TCAGTACTTC	16980
ACCGTTCCTA	TTGCTGTAA	ACGTTGCAGT	ATAATTATTG	GTTGTAtCTC	GGAATTTAAA	17040
TTGTGCGCCT	GCTAAACGTT	GTTTCGTATA	TCCATCTTGC	TTAACAATTT	TGATGGCGCC	17100
TTTTTGAGCT	TGATTATTTT	TGGTGAACGT	GTAGGTTTTG	CCAGCATAAT	TCGCAGGTAA	17160
AAGAATTTTA	TGCACCGTTC	TGTCAAGTAT	ATAACCATTT	GGTGCTTTTA	TTTCTGTAAT	17220
TTCATAAATC	CGGTTGACGG	AAAAATTATA	CAAAGTTGCT	TCACCATTAG	CATTAGCAGT	17280
TACTACGCCT	GTATTCCCAT	TCACTGTGTC	ACGCCATCTA	AATTGTGCGC	CTGCTAAACG	17340
TTGTTTTGaG	TAACCATCTT	GTTTAATTAA	TTTTAAATTG	CCACGTAGCG	CAACATTTTC	17400
AATGGTATAA	TAACCAATCT	TGTCAGCAAA	ATTAGTCGTT	AGGGTAACTC	GGTAAGTCTT	17460
GTTATTTAAG	ATATAGCCGT	TTGGTGCTTT	AATTTCTGTA	ATTTCTGATA	CACGATTGAC	17520
TGGCAAATTC	GTTATCAGGG	CTTGTCCGTT	TTTGTACGCT	ACAACGTGTC	CTTCTTTTCC	17580

TGTCACCGTA	TCACGCCATT	TGAACTCTGC	ACCTGCTAGA	CGTTTTCTTG	AATCTTTGTC	17640
TACTTTGATT	ATTGCTAGAC	TACCTTTTTG	CGCTGTATTT	TCAATGGTAT	AATAACCAAT	17700
TTTATCwGCA	AAATTAGTCG	TTAAGATAAC	TCGGTAAGTC	TTGTTATTTA	AGATATAGCC	17760
rTTTGrTGCT	TTAATTTCyG	TAATTTTCGTA	TACACGATTG	ACTGGCAAAT	ctGTTATCAG	17820
kGCTTGwCCr	TTTTTGTCAG	CTACGACTGT	TCCTTCTTTT	CCTGTCACCG	TATCACGCCA	17880
TTTAAACTCC	GCACCTGCTA	GACGTTTTCT	TGAATCTTTG	TCTACTTTGA	TTATTGCTAG	17940
ACTACCTTTT	TGCGCTGTAT	TTTCAATGAT	TACTGTCACT	ACTTTATTAG	CTTGTGCAGT	18000
TGTTAGAGTG	ACTTTATGCA	CAGTTTTATC	TAATACATAC	CCCACTGGTG	CTTCGTTTC	18060
TGTGATTTCa	TACGTACGAT	TGACTGATAA	GTTGGGAATT	GTGATGGTTC	CATCGGTTCC	18120
AACTGTCACT	rTTCCTGTTT	TTCCTGTCAC	TGTATCyTTC	CACTGAAATT	CTGCGCCTGT	18180
TAAACGTTTT	TTTCTkTCTT	TGTCTTGTTT	AACAATTTTG	ATGGAGCCTT	TTTGCGCTGT	18240
ATTTTCAATG	ATTACTGTCA	CTACTTTATT	TGCCTGTGCA	GTTGTAAAG	TGACTTTATG	18300
CACGGTTTTG	TCTAGTACAT	AACCTGTGGG	TGCTTTTGTT	TCTGTAAGTT	CATACGTACG	18360
ATTAAGTGCT	AAATTGGGAA	TTGTGACCGT	TCCGTGCGTT	CCAAGTGTCA	CTGTTCCCTGT	18420
TTTTCTGTGC	ACTGTATCCT	TCCACTGAAA	TTCTGCGCCT	GTTAAACGTT	TTTTGCTTTC	18480
TTTGTCTTGT	TTAACAATTT	TGATGGAGCC	AATTTGTACT	TGCTGCTTAT	TCTTCAACCG	18540
ATATGTTAAA	GTGTAGTCGG	CTTTTCCAGC	AAATGAAACA	GGATGTTTAG	TTTCGTCTAA	18600
TTCATAGCCA	TTAGGCGCCT	TCGTTTCAAC	TAAATAATAT	TCTCCATCAG	GTGATAAATT	18660
TGAAATAAAC	AGCTCGCCT	TTTCATTCTG	GACACCTTTT	TCTATTAGAT	TATTTTTTkG	18720
GTCATACAAC	GAAAATTTCAG	CACCTGcTAA	ACGTTTCACC	TCATTATTTT	CATCTACTTT	18780
TTCyACTTTT	AACTGGkTTn	CTAAACGCGt	ATTTCTTTTt	tCAAGCGAAT	CGTCTGkGTT	18840
TGTTCTAcTT	AATTnCAAAT	GGAaCTGGTG	TTGCATCTAA	CTTGTAACCA	GTTGGTGCTT	18900
TTGTTTCAAT	GAATGCATAC	TGACCAGCTT	CCAAACCAGC	TAAACGAATT	TGACCTGCTT	18960
GATTGGTGGT	TAATGTTTCT	TGAACAACTT	GTTGTGTTTT	TGTGTTACAG	ACTTTGAAAA	19020
CAGCACCTTC	TAAAACTTTA	GcAGTTACTG	CAcAATTTTT	TCyAAAaCGA	CACCCCyTCG	19080
ACCTGTTCCA	TCACTACTCT	CGTCTGCTGT	GATGTAGGCA	ATATCATCAT	CTTCCCAaGC	19140
ATCGAATCCC	TCACTAGAAA	TACGCATATA	CGTGTAATTT	GATAGCACGG	CATCACTTGG	19200
AATATTTTCA	TTTGAATCG	TTTCTAAATA	AATATAATAG	GCATTTTTAT	TGGAAAGATA	19260
ACTGCCCGTG	GCTGTTGGTA	TTCTTCAAAA	TTTCTCTGCT	TCATAATTAA	ATCCAGATTG	19320
AGAGGCATTT	TTTTGGAAAA	ATTGACTGGC	ATCATAGCTA	TATTCGTCTAT	GTTCCAAATT	19380
CCGAATCATT	TCACCATTTG	TTTCTAAAAT	AGGCTGCTTG	GTGAGATATG	CAGCAATACT	19440
GTCAATAGCT	AATGTTTGGT	TTGGACCTAC	ATTATCTAAA	AATTTAAAGG	TTGTTTTtTTC	19500
TGAACCTGAA	AGACCTGCTT	GATTATAGTT	AAAATAAATG	ACCCAAGCAA	TTCTTCCCTC	19560

606

TGAGGTAAAT CCTAACGGGA CTTTTCGAAC CAGTTCAGTT GGCGGAACTT TCTCCACGTA	19620
TTGCCCCGATA TTAATTTTGA CAATTTTGTG TGAATCCCCA-AAAGTAAAAA CGACTTGATC	19680
TTCAAAATCT TCTGATAAAT TTTGATTATA TCGTAGCTTC AGTTGCAAAT CTGCTGAGTA	19740
ATCTGTAAGC GTTTGTAAGG CTTTAATCGT CCAAGTACCA TCATCGTTTG GAATCACTTG	19800
AAGTGCGTCT CCAATTTCAA AAGGTGTCAT TGAATAGGAA TATTTTTCGC TAGCCATCCT	19860
TAATGTCATC GTCTCTCCGC TTGCTAATTC AGAGATTCTT AATGACAAAG AAAGTGTGCGC	19920
CCAATCTTTT GTAGCAACAA TTTCCGTTTT TTCCCCCGGn yATTCAACAG CTATAGACCG	19980
GATTTGTGTG CCACTTTCAT TATCCATTGC ATAAGAAATA ATACTTGTAT TAGATCCTAT	20040
AATAGTAAAG ATAATTAAAC AAATCGAAAA TA	20072

## (2) INFORMATION FOR SEQ ID NO: 90:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 688 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

TTGAnACAGA TGGAGTnnGC ACTGGCTATA CAGTAAGATA TGATCTTGGT GGAGAAAAAG	60
AnAAAGACAT AACAGTCAAT TTACCCGGAG CAGaATTACC ACTTAGTTTA AAAGTAAAG	120
CACATTACAC AAATGGATTA CCACTTGAAG CGGGAAAAGA AGTACGCTTA AGAAATTTAA	180
CAGATGGATC AACAGAAGTA mTTAAGAAAC ArGTAGACGC TAACGGAGAA GTTGTATTTA	240
CTGAGCAAGA TGGTATAAAA AAGGAAGTAA ATTATGGTAT TGAAACAGAT GGAGTAAGAA	300
ATGGCTATAC AGTAAGATAT GATCTTGGTG GAGAAAAAGA AAAAGACATA ACAGTCAATT	360
TACCCGGAGC AGAATTACCA CTTAGTTTAA AAGTAAAAGC ACATTACACA AATGGATTAC	420
CACTTGAAGC GGGAAAAGAA GTACGCTTAA GAAATTTAAC AGATGGATCA ACAGAAGTAC	480
TTAAGAAACA AGTAGACGCT AACGGAGAAG TTGTATTTAC TGAGCAAGAT GGTATAAAAA	540
AGGAAGTAAA TTATGGTATT GAAACAGATG GAGTAAGAAA TGGCTATACA GTAAGATATG	600
ATCTTGGTGG AGAAAAAGAA AAAGACATAA CAGTCAATTT ACCCGGAGCA GAATTACCAC	660
TTAGTTTAAA AGTAAAAGCA CATTACAC	688

## (2) INFORMATION FOR SEQ ID NO: 91:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10194 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

nGATAAGGTC CTATAAATAG nCCGGATATn TTTTATTAT TAAGATCTAG TACTTTCCTC	60
--	----

TTACTATGAG TGTACTAGAT TCTTTTTTTG CCAACTTTAA AGAACTAGAC CAATTTTCGG	120
TTTTTAGATT TTTGTTGACA AGTACCTACT TTTAATTGTA TGATTGTTTA TGTTTCTACG	180
GAATATATTT ACAAAGGAGG TATATAGACC AATATCGGAT AGCGCCAGAC CTGAACGTTT	240
AGGTGACGAG GAGAGAGCTT ATCGAAGATT CGGCGGGTGG CTCTAGGGAC TGCACTCTAC	300
AGATAACAAA GAAAACTAA TTGTGAAGTT AGAACAAAGC GGTATCACG CAGGTAGAAA	360
CATATAATGG GATTTGCTAG TTAAACAGGA CCTAGTAAAT ACAGAATTTT AATTGGAAAG	420
TAGGAATTAG CATTATGTGT GGAATTGTCG GAATGATTGG AAAAGAAAAT GTCACGGAAA	480
GTTTGATTCA AGGATTAACA AAATTGGAAT ATAGAGGTTA TGATTCTGCT GGGATTGTG	540
TGAATGrTCm AAGCCAACAG CCCCATGTAG TTAArGCAGT TtGGGGgCAT TkCCsATTTa	600
GCAGCAAAAT TAACGCCAGA AATCGATgGA ACAGTTGGTA TTGGTnCATA CTCGTTGGGC	660
CACTCATGGC GAACCAACTG TCGCCAATGC ACATCCCAT GTTTCAGCG ATCAACGCTT	720
TGCATTAGTG CATAATGGTG TCATTGAAAA CTTTGAAGAA TTAAAAGAAC AATTTTAAAG	780
TGGTGCTCAC TTAATTGGTG ATACAGaTAC TGAAATCGTC GTTCAATTAA TTGcTCATTT	840
TGCTGAAAAT GGTTTTTCAA CAAAAGATGn CTTTTAAAGC GGCTTTAGCm GAwATTaAAG	900
GTTCATACGC TTTTGCATTA ATGGATAAAA TGGaCCCCAA TATTATCTAT GTAGCGAAAA	960
ATAAAAGCCC GTTACTTATC GGTCTAgGTG mAGGCTaAT GkTTATtTGT AGTGaTGCGA	1020
TGGcGATGAT TAAAGAAcCA ATCAATTtGT TGAGATtGTT GATGGTGAGA TTGTAAGTGT	1080
TACAGCGGAC AAAGTAGTAA TCGAAACGCC AGAAGGACAA GTAATCAACC GAGAACCTTT	1140
TGAAGCGCAA TTAGATTTAA ATGACATTGA AAAAGGAACA TATCCGTTTT ATATGCTCAA	1200
AGAAATTGAT GAACAACCTG CAGTGATGCG CAAAATTGTT CAAGCGTATA CTCAAGGTGA	1260
AAAAGTGA CT TTAGATGAGG CGTTACTGGA AAAAATTAAT CAATGTAACC GAATTTACAT	1320
TATTGCTTGT GGGACAAGTT ATAATGCAGG TTGGGTCGGA AAATCTTTAA TTGAAACGAT	1380
TGCCAGTATT CCTGTGGAAG TTCATTTATC AAGCGAATTT GGCTACAACA TGCCTATTTT	1440
ATCTGAAAAA CCGTTCTTCA TTTTCTTAAG TCAAAGTGGT GAAACGGCAG ATAGTCGCCA	1500
AGTACTGGTA CAGGTTAATG AATTAGGGCA TCCATCACTT ACATTAACAA ATGTTGCTGG	1560
CTCAACATTA TCACGTGAAG CGGACGATAC CTTGTTATTA CACGCTGGAC CAGAAATTGC	1620
TGTCGCTTCC ACCAAAGCTT ATACTGCTCA AATCGCTGTC TTAGCTATTT TAGCAAAAGC	1680
AATTGGCGAA GAACGCCAAA CATTAGCAGC CGTCTCATTT GATGTGGCTC ACGAATTGAG	1740
CGTGGTAGCA GCTGTCATGG AAACCTTGAT TGATGAAAAA CAACAAATGA AAGCCTTGGT	1800
GGAAGATTAT CTGACGTATA CTCGAAATGC CTTCTATATC GGGCGAGGAG TCGACTATTA	1860
TGTGTCGATG GAAGCGGCCT TGAAGTTGAA AGAAATTTCC TACATTCAAG CAGAAGGCTT	1920
TGCGGCAGGT GAATTGAAGC ATGGAACGAT TGCCTTAATT GAAGAAGGAA CACCTGTGAT	1980
TGGGATTGTG ACAGATGCCA AAGTCGCTGC CCATACAAGA GGCAATTTAA AGGAAGTCGA	2040

AAGTCGTAGA	GCAAGAAATA	TTGTGATTGC	TTTGGAAATCC	TTGGCAAAAC	CACAGGACCA	2100
ACTGATTATT	CCAGATGTTT	ATCCGTTACT	TTCTGCCTTA	GTTAGCATTG	TTCCAGGTCA	2160
ATTATTAGCG	TACTATGCAA	CCTTACAAAG	AGGCTACGAT	GTTGATAAAC	CAAGGAACCT	2220
AGCCAAAAGT	GTAcTGTGTA	ATAAATCAAG	AACCTGAGAG	AAAGATGTTT	AACGGGTTTT	2280
CATGAGTTTT	AATATCTGAG	ACAAAAATCA	CTTTAGATTT	TTGTCTCAGA	TATTTTTTAT	2340
AAAAGAAAGA	TATTCAGTCA	AAAATGAAAG	ATTCCAAAGA	AATTCTTTTT	TTCCGTTGAA	2400
AACTGTTTGC	ATTTTAGCAA	AATTAGATAT	AATTAATAAG	AACCGAaaa	TAAGAATTAG	2460
CCAGTGAGGT	AACTTTGTTT	GAAAGAAAGC	AAAAAGTAGG	TTAAGCTGAA	AAGTGCTAAT	2520
TTATTTAACA	AAAAAAGATT	GTGAGTTGTT	TGATATGGAT	CAGTTTAGTA	CATTAAGTGA	2580
ACAAGAATTT	ACTACCTTTG	CAATGACACA	TCCTGCAGGT	AACTTTTTAG	AAACACCCGA	2640
AATGAAACAT	TTATTGGAAC	GACGTGGCTG	GCATTGTGAA	TACGTTGGCG	TCAAACGAGA	2700
AGGCCAACTA	ATTGCTGCGT	GTATCCTAAG	CAAGAAAAAA	GTCAAATTG	GTTATGCTTT	2760
CGACATTGAC	GGCGGAATTT	TGATGGATTA	TACAGATAAA	AAATGTGTAG	AAGCCTTTTT	2820
CACAGGCTTA	AAAAAATATG	TGAAAAAA	TGATGGCCTT	TATTTAACCT	TTACTCCAAA	2880
TAAACAAATT	TGTTTACGTG	ATTTTAATGG	CGGTGAAGTT	GAAAAGGTCA	ATCAAGAAAC	2940
ATTTGACTAT	TTCACATCAA	TTGGTTTTGA	ACACCAAGGG	TTTGATGTGC	ACAATTTTGA	3000
TGGTGCGCCA	CGTTGGCTGT	TTGTCAAAGA	TATGGCTGGA	TTAACAGAAG	AAGATCTTTG	3060
GAAATCTTAT	GGAAAAGATG	CTAAATATGA	CATCAAGAAA	ACATGGGAAT	ATGGAGTAAC	3120
AATCCGAGAA	CTTCgTTAcG	AAGAGCTACC	TCTCTTTAAG	AAATTAACGG	AAGAAACAAG	3180
TGCACGGCGT	AATTTtGAaG	ATAaGGATTt	AGCTTATTAC	CAaGCTGTTT	ATGAaGAGTT	3240
TGgCGAaCGA	gCGAaTTTAT	GGTGcGGACT	AAcTTTGCTA	CTTATTTaGA	AAATCTTCAT	3300
GAGAAATTAC	GCAAATTACA	AGAGACGCTC	AACGAAgTCA	ATGAAGCGTT	AATTGCCAAT	3360
CCTAAAAGCC	GTAAAAAA	TAACCAGAAA	CGTGAATTTG	AAGATGAAGT	TCGTACCGTG	3420
CGCAAACGGA	TTGACGAGGC	TAAAGAAATG	AAAACATCAG	ATGAACCTGA	AATTTTAGCA	3480
GGTGCCTTAT	TCATTGTTCA	TCCACAAGAA	GTTGTTTACT	TGTTTAGTGG	TACGTATGAA	3540
AAATATAAAC	AATACTATGC	ACCGTATTTA	ATTCAACATA	AAATGTTAAC	TTACACAGTA	3600
GAAAATAACA	TTCCCAAATA	TAATTTTTAT	GGCGTTGACG	GTATCTTTGA	TGGAAGTGAT	3660
GGGGTTCTTA	AATTCAAACA	ATCCTTTGGC	GGTCATGTGC	AAGAATTAAT	GGGCAATTTT	3720
CAATGGAAAG	CCAAACCAAT	GAAATATGCC	TTATATCACG	CATTAAAAAC	AATTAAAGAA	3780
AAAGTCTAAA	ATTAAAACCC	AATCCTTAAA	AAGGAATGGC	AAAGTGAGGC	GTTTAAATGG	3840
AAAATCAAGT	CGAGGTTATG	ACGTATGCAC	ATTTAAAAGA	AATTATGCAG	GCGTTAGAAG	3900
CAAACGAAGC	GATTACAGAA	GATACAAAAG	TCTTTATTGA	TACAGGATGG	GATAGCGTGC	3960
AAGAAGTAGC	CCCAGATGCG	GTCAGTATTG	AAAAAGTTGC	TAAATTTACA	GTAGCAGATG	4020



TTTAAACAAA	CGAAAGTTTT	GCTGGGTATA	GTTTAGAAGA	AAAGGCGGAG	AAGATGAACG	4080
CTGAGGGCGA	CCTAGAAACG	GCAATTATTA	TTCGTAATCT	TTACTAAAAA	AGTCCGTTTG	4140
AATAGTTTAA	ATCGAGGTGG	GGCAGAAACC	GTTTAGCTCC	GAGCAACGCG	TAsTGTGGT	4200
TCTGTTCCCG	CGGTTTATCA	GSTTTTGAGC	CTGGAGCAAA	AATCCAAAGT	GATTTTTGTC	4260
CCAGGCTCAT	CTGCTTTTAA	TAGAAATAA	AGGAGGTTAG	ATATGAGTAC	CATTCTmTGrA	4320
TTYCCAAAAA	ACTATGAGCG	TTTTATTGCA	CAAGGAGAAG	AAGCACTTGT	AGAACATAAT	4380
CAGATCGCCG	CTTTAGAAAA	TTTTCAACAA	GCCTATCAGC	wACCAmCAAC	GCCGCCAGTA	4440
AACCAAAAAA	TCTGCCAATT	ATTGTTGGAA	ATGGGtGAaG	CTGATGAAGC	ACTAGCATTG	4500
GCAGAAGCGT	TCCAGAGATC	TTATTTTGAG	AACCTTGAAA	CAGCGGCAAT	TTATATGCAA	4560
ATTATASTTC	AAAGTCGTGG	CTTTATTGAA	GGGTATATTT	TATTGAAGCA	ATTACTTCAA	4620
ACTLAAPAAA	TACGCTAGC	ACACAAAAAA	ACGTTAGAAC	AGCAATTAAT	GCAAGTGGAA	4680
GAGGCGTATC	ACCAATTAGA	AACGCAGCAA	ATTCAAGCAA	TTAAAAGGrA	CTTGTTAGTA	4740
AGCGATCAAT	TGCCCCGTTA	TCAACAATTA	GCGAACATTA	AAACCAGTTT	ATATTTGCCT	4800
AAACCAGTAT	TCTGGGAGT	CGCAAAAGAT	TTAGTTATGA	ACCAAGCACT	TAGTTATTTT	4860
GCTCGGGAAAT	GSTTCATTGA	AGAATTAGCA	CTTTTGCAAT	TTTCAGAACC	ACTGACTTTT	4920
CTTGGGTATG	ACCAATCGCC	TGAGACAGTC	TTACTAGAAG	GAAAAACTGG	ACCGTTAAAT	4980
ACGCCGATCT	ATTCCGAAAT	TTGTACTGAA	CTTAGGAATC	GTTTAGAAAA	TGATGATCCA	5040
ATTATGCTGC	AGCATTTCGA	AGAAGAAATT	CGATTGCATC	TTGCCTATTT	GTATCCATTA	5100
GCAGAGACGG	TTATTTTCAG	TCCAACATC	TGGGTCTTGG	GTTATTTGGC	AACCTATTAT	5160
CCTGAATATA	TTGAAAPAGA	ATTGACTGAA	GCAAAAGGCC	ATCAAATCGA	TGCTGTTCAA	5220
AAAGTTCAAC	AGGCGATACG	AACGGCATT	ACACAAATAA	TGCTTTAAAA	AATGGCACGT	5280
CCTAAAGCG	AGGACGTGCC	ATTTTTTCGG	TTTCAGAAAT	TTGCTGAATC	TGCTCTAATT	5340
GCTTTTCTTG	ACTGATTATT	GCGGATAATA	GAAAAAAGCA	AGACAGAAAT	AAGTAAAAAG	5400
AATTCCTTGA	ATCAGGCPAA	TCACAACGAA	CAATGGCGTT	GTCCTGGAA	AAGAATCAGA	5460
AATCCGTGTA	ATTAAAPATT	CACTAACCA	TAAATCATT	GAAAATAAAA	ATAATTTTTT	5520
CTGACGAGTT	CTCATAGTCC	GCAACAACCT	TTCTGGTGAT	TTTCACTTCT	CTATCATAAC	5580
ATAGAAAPACG	AGCTGTTTTT	TCTCAAAAAT	TTTGTTACTT	CATAAGAAAA	ATGGGAAATA	5640
AGCAAGTTTC	AGGAGTACT	AGCTTTACCG	AAAAGAATAC	TTAATCCATC	TTAATCAAAC	5700
CATACAGTCA	CSTTTATTCG	AAAAGAAGTG	TTTTTGATG	GAACAAATTA	TTTCATAAAC	5760
ACAAATATTT	CTTTGTTGTE	TGTATAAATG	TCATAAAAGT	GATTACTTAC	TACTTTTTGA	5820
TTTTTAGACA	TATCCGTATC	GGTCAGTGGT	GCCATCTCTA	ATCCTAAATT	CGATGAAGTA	5880
TTGAACCAAA	CGTAATTAGG	GAAGTATAGA	CCGTCATTTG	GTGGCACAAT	TTTTGAAAGA	5940
ATTGGGTAGT	TTCGACTTGT	ATTAGCATAA	ACAGTTGAAT	TTTTGAAAAA	CGTATTTAAA	6000

TAACTTTTT TTCGTTTCGTT TGTTACAACA TTTTTTAAAT CATCTGTTAA AACAGCTGCT	6060
TGGCGATTGA ATGCTTCTTT TTGATAGTTT AATGTGGAAG AATACACCAA AACAAACGAT	6120
AAAATATAAT AAGATAACAA GCTGGCCACA AGTAACGAAG TTAGATAAAT TGCTTTCCT	6180
TGTCGATTGT TAAATAAACT TAACATGCTG ATTGTAACGA AAACAGCAAA TCCGTAAATA	6240
TAGCGAGGCG CATCACCTGC GATATTTCTT GAATAGACAA TAAAAATGCC AAACTTAAG	6300
AGAGAAGCAA GTCCTAAATA CAAGATACAA TAAAGGAAAC TTAAGCCTAA GTTGATTTTG	6360
CTATTTGATA GATGACTAAT AAAAAATAAA ACAAGCAGTA GGAGCAGTAA TAAAAGCCAT	6420
AAACGATTAC TCTGGTTCAA TAGCTCTGAA AAATAATGGT AACTATTTTG ATAAGCCACA	6480
CTGGGAATAT CGTGCAAAC TGGTAAGGCC TGACTGCCCC CAATGTTGTT TGTTCGGAA	6540
GGTACAAATT TTAACTCGAT AAGGTAGCTG ACTATAGAAA GTAGATAAGC GATAGCAGCA	6600
AGTGCGACTT TTTTGCAAAG AGCAATGAAA TTTTCTCCAG CAAGTAACTG CTTGAAAGTA	6660
AGGAATAAGA CGATAACTAT ATAAATTCCT GACGAGGCTT GGTAAGTATT AACATAACA	6720
AACAGTGAAA AACAGAAAC CAAGAAAAAC GTAAAAGAAT TACGCTGCCA CCAATAAAAA	6780
GGAAGAAACG AACAAAAAAT ACTCAATGCC ATATAGGGAC TATCAAACCT GAACTAAGA	6840
CATTGGAGGA ACCAAGGATT GAGGCCAATT AAAGTAGCTA CTAAAGAGGA ACCCCAGCTA	6900
ATTTTTTGGT CATTGATGAT ATAAATTGCC aGTATGCTTG TGAGAGATAA GATAAGACCT	6960
GTTAGAATCG GTGTGGTCAA GCCCAGATCT GTCAAATGAC GACTTCCTTG AACAAACCAT	7020
GAGAGAAATT CACTGCCCCA ACGTGCGTAA GTCGTACCAA AATCTGTCAT TCCAGCGACT	7080
TGGCGAGCAT TATCATCTAA GTAAGGGAAA TTGATAATGC CGATGGATAA AACTGCCACT	7140
TGATAAATTA AAATGATCAA TAGAAGATAG GCTTTATTTT CAATACACCA ATTCTTTAAA	7200
TCAAATTTTT TTGATTCKGT AGGTAATTGT TTCATCATTG TACTCCACTC CAATAATTAT	7260
TTAmCaATTA AAATATmCCA TAACGAGTGG AATGAAGAAA TTGCTTTTGT TATTTTCTCA	7320
TCTTGACAGA AACTAAGGG TAGGCAAAAC AAAAAATCC CAACTCTTTA TTAAAGAGTT	7380
GGGATTTTAC CAAAAGCCA CCTGCCAGCT TCATAGATAA TTATTCAAGA ATATTGCTGT	7440
TTAATTAAAA ACAATTCTAT ATGTTTAAAT ATATACATCA AGATAACAAT TTAGATATAC	7500
TTTGGGTAAT TTTTGGGTAT TTACAAAATA TCTTTGTATT TGTTTTCATA TTTTCTTTG	7560
TyCTTTTTTC TATTTGCTTC AGTAGCTGAT GCGTATTCAT ATGACTGTAT TATCTCTTTA	7620
TGTCCTAGCT GTGTTTGTaT ATATATTGGA TCAGCTTCAG CTAGGaTTAA TAAGTCAGAA	7680
CGGGcATGTC TAAACTATG AGATGAAATT CtTTCAATAT TTGcTAGTTC ACATGTTTCT	7740
TTTATATATT TATTCACAGA TGAATTTGAG ATGGGCCATG AATTTTTTTT ACTAACAAAA	7800
GCAAATTTTT GGGGATTACG TAAATCTCCT CTCAATATTT TTTTCTTTTG TTCCATTTCA	7860
TATTTTTTTA AATAATAGAA AAGAGAAGAA GGGACATCAA TATTTCGAAC ACCTGCTTCT	7920
GATTTGTAG GGCCAATTCT ATACTGGCCA GCATCATACT TAtAGGCATT TTTATCTTA	7980

ATAACTTCGT TACTTTCATC AATGCTATCC CAAGTCAACG CAGCAACCTC TTGATAACGA	8040
GCTCCAGTAT ATAGTTGTAC TAAAATGAAG TATTTACTTG TGGTGTTC ACGAGATAAT	8100
AAAGCTTCTT TAAATTTAGT TAAATCATCA ATAGATATGG TTTTATTTGA TTTCTTTTCT	8160
GGTACTTGGC CAGTTAACTT TATATTCTT GTTGGATTG AATGAATATA TTGCATGTCC	8220
AACGCTTCTA AAAACATAGA TTTTACAATG TTATGTCTGT TGCTCACCGT CTCAACTGCT	8280
AGTCCTTTTT CATTGATATT GCCAGGTTTT GCAAGCCAGT TAATCCATTG CTGATATTTG	8340
ACTTTGGTGA CGCCTTTAG CAATTGTTCG TCCCCAAAGA AATCATGTAA ATTCTTTCTT	8400
GTTAAAGCTA TTTTTGTCAT ACGGCCAGCA GACACTTTAT TCTTTTTAAA TGTATCAACC	8460
CAAATATCAT AAAATCTAA AAGTTTAATT GCTTCAGTGT TTTCACCACG CATAATTTTT	8520
GATTTAAACT CTGTTTCAGC ATCTCTACAA GCCTTCTTAG TAGATCTAGT AATCATTTTA	8580
TTTTTCCATT CATTTGTAAT AGAGTCTTTA TATTTAAATC TTAAGTCCA TTTTCCATTA	8640
GCTAATTTCT TTGGTTCTGC CATAAATACC AACCTCTCGG ATTTTTTTAA TTTAAATACA	8700
TTTAAGAACG ATTTTTTATT GAATATGTTA TAAGTAAGTA TCAAGTTATC GTATATCATA	8760
AATACGAATA CGAATATATG TtCGTTTTTT GTGaTAAAAA GAAAAACCCG AAGGTTTCTC	8820
TCTTTGTAA TAAATAATA AGAAGATATT AATAATCAAT AATTTCTACA AGTCCGCTAT	8880
CTGCACTTTG TGGAGTTATG TGTATTTCCA CATTAGCTTC TCTTTTGCT CCATATGCAT	8940
TTACTATAGT AGCACTTACT TTTTGATACC ATTTACCATC TTTTGCGTC CATGGTTGGg	9000
ATGACGCCGA TAATTGAATG CATTTTGCTT CCTTTATAAG GATACATATC ATTAAATTGA	9060
TGCTTAGCTA AAGTATCTAA GACGGTTTGT TGTTCTAAGG TTGGCTCATC TGCTAGCTTA	9120
GTAATGTCAT TAGTATTTT ATCTTCGCTC GGTAAAGTAT TAGTAGCTAC ATCAGACTTC	9180
GCTTCAGTAC TTAATTCTTT TTTTCAGAT ACATCAGATT GTTTATTATC TTGGCTACTA	9240
CTTTTAGAAA CATCAGAGTT TACTGTTTTk TGTTTGCTAG TTTCAGTACT kGATACTGGa	9300
TTTTTCaTat ATTCTTGAAA ATCTTTTCA TTTAAAAATC TAACAACAAA ATCTTTGTCG	9360
GCAGTTTCGA ATGTGAGAGT ATCTCCCTTT TTAAGAGATA CTTTTTCTT GTTTCCCACT	9420
CCGTATAAGT CATGGTTACT ATTTTCTCCG CCTGATTGA CGGTAATAAG AACTATGCTC	9480
GATTTATCAT CTAGTTGAGT TAATACAGCA TAGTAATCTC CAGGATCAAT ATCTTTTCCA	9540
aCTGAATATT CCCCTGACGC AACAAAGtCT TTTGAATTCT GTTGTTTAAAC TGTGCTGCTA	9600
GATTGACTGG TTGACTTTTC AGTATCTTTA TTATCATTAT TTCCCGrACA AGCAGTAAGT	9660
AGTGAAATAC CACAAACACA AACCAATCCT AAATAATTA ATTTTTTCAT CATAATTCCT	9720
CCTAAATAAT TTATTCAAAC CATTCATATA CCCCAATTGT ACCATCACAC CTAAAAATTA	9780
TTTGATAATT TTTATAGATC ATTCCATTTT CATATAATTT TGAGTAACTT TTGATTGCGT	9840
CTTGAAGATA TTCAGTAGTT ACTCCTAGGA ACTCTGCACA TTGCCAAACA TATATTAATC	9900
GCTCATGATA ACAATCTATA AAATCTTGTG GAGTTATTAA CAATTTACAT CCACAATCTC	9960

TAGCTCGTTG TTCTGTGTG CTTTTTTCAT TACTATTTTG CTCTATAATA TCCCCAACAC	10020
TTGTTAAATG ATGACCAATT TCCTCTGCGA GAGTACCCAC CATTCTTCG TTTGATTGAT	10080
TAGGGkTTAA ATATACTACA TTATTAATAT AGAGACCTTT CTGATTTCcTG GGCATCATAG	10140
GTCTATTCTG TAGTTTAGTT CAGGATAATT AGCCATCAAG CTTTCCAATT TATn	10194

## (2) INFORMATION FOR SEQ ID NO: 92:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2053 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

nATTGGTTGG TACAGGTATT nGAATTGGTT nGCTCCCTTG GCATCAAAAG GTTTGCmAAT	60
tCGtTAAwTT mTyCAGATTG GTAATATCTG aTAACTCGkT GtAAGAAcAA TCaATAGAAA	120
aTAGATTGTG ATTTTgGCTG CTAACAtCTA GAACAGTTAA TTGATTGCTC aAACAAACTA	180
AAATGGCtAG ATCAACATTA TTAGTAACAT CTAaAACGGT TAATTGATTG TAAGAACAGT	240
CTAAAGATG TAACCTAGTA TTTTGGCTGA CATCTAAAGT AGTTAATTGA TTGCTAGAAC	300
AGTCTAAATT CCTTAAATAG GTGTTCTGAG TAATATCCAA GTCTGTTAGA CGATTTCCAC	360
TACAAGATAG AGTCTCAAGC ATTCTATTCT GACTAAATC CAGATTGGTC AAATTATTAA	420
ATTGACAGCT TAAATCAGTT AATTGTTCTG CATACTTGAT ACCTTTCAAG CTATCTAAAT	480
TTTTATTAGT GACAGTAATT TTTGTTAAAC TTGACATATC TGCAATAGTC AATTGATCCT	540
TATCTTTCCC AAAGCCTGAC CAGTTATTTT CTTGGCTTTG TACAGTAGAA TCAGTCCTTA	600
AAAGTGTTC GAACCTAGCA TCGATATCTG TTCCTTCAAT TAAATTGGTT GTTTTAGGAT	660
CTAATCTTTT ATTTTtTTG TCCAGTTTTG CCGATTCTGTT TTCTTGAGTA GGAGTAGTCT	720
TTGTAGATAA AGTTGAATTT AATGTGCTCG TTGTTGTAA TTCCATTGAG TTTTTGTTC	780
CTTGTGATTG ATCCGTATTT GCATCTATAG AGGAATCTGT GCCCgAAAGT AACTTGAAG	840
ATGATGTTGA TTCAACTTTT TTGGATGAAG TAGTTTTTGC CGATACCATC ACACCCATTA	900
TTCCTTGTGA AAGGACTAAA CCTAAACACA AAAACAATCT AAAGGATTTA CTCATAGTTA	960
TTCTCCTTTT CTACATAATT GTCTAAATAG AAACAGCCAA AACTACTTA TTAAAAACAA	1020
TATGCCAAAC CAGCTTAGCC AAGTTATTCG ATTGTCACCA GTTTTAGGTA ATCTTTTAT	1080
TTGCGTTCTG CTACGAACGA TTGATAGCGA TTGAAtGGTT GTTTGCGGCA ATACATtCAC	1140
AACTGGATTT TTTTTGATG GTATGTCTGA AGTTTTGTGA AACCCGATTC CGACTTCTGT	1200
GGTCTCCCTA TTTTCTGCAT AAGTTGTTT AGGAATATAA ATTAATAGAA CCAaCAaTAA	1260
TAAAAaCkGA AGATTTCyTC yTCTCATTAT GTCACTCCC CTTCGTTTTT CCACTTCTCT	1320
CCACGACGAA CTAAGTAAGC AATCAATCCA GCAAGTGCTA CTGTCAAAGA TAATGCTACC	1380
CATGGTACCC ATTGTGGGAT CACTACTTGA TACGCTGCCG ATTGATTCAC TTTTTTCGCT	1440

TGCGCTTCTC CTACTTTAAA TTCTTTTCC CATTTCGAAG TTCGACCATC ACCTGTTGCT	1500
TTTCCAGTAA AGTGATAAGT TCCTGGTTTA AACGTATTCA ATCCAATAGG AATGTTAAAA	1560
TCGAAATTCG AATTTGGTGC TACAGAGAAT TTCTCCATCT TATGTTTATA TAAAATTTTC	1620
TTTTCCCTT TTAGACGTAT CTCCCCTGA ATCGTTAGTT CTTGAAGAAC CTTTGGTTGA	1680
TCGTTTTGAA TACGTGCTGC AATCACTTTT TGACCATTCTG CTAATTGTAA GCTGGCTTTC	1740
TTTAAATGAA GCATCGCTCC CTGATTAAAG GGTTCCTCAT CCTCCGTAAG CATAATCGCA	1800
ATCACGCGCG CATATTGAGG AGTTAAGCCA GATTTCTTTT GTTCACTTC CGTACTATTC	1860
TTTCGAACAA AGCGCAGAGA GCCCAGCTTT ACTCCGGAAG AAGCCGATTC AGGTGGCGTG	1920
GATCTCAAAT GTATCGGTT TCTCTTCTT CCCCTTGGAG GGGGAATTTT CTTAACGGCC	1980
ACCTTTnATT TTTTACTAAC TCACGGAATG GGGGCTTTCA AACGGGCATC GAACTTAnGG	2040
GGAACATTG GnA	2053

## (2) INFORMATION FOR SEQ ID NO: 93:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4178 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

GCAGTCGAGC CTGTTGGTCC ATAAAAATCG GTGTATGAAA GCAATTCCTT TTAGATATTT	60
TTGACTTCCC AATTTGATAT ACTGCGGAAA GTATCTGAGA ATTTATTTAA CACATTTTTT	120
ATTTACTAnA GAAAGTAAAG ATCaGAGCTG TGCaAATAGA AATGATTAAA AACATCCCaA	180
TATCATCATC TAATGTTTCA AAACTTCAA CAGGTTTCAT AAAATGTTA ATTATTTTTT	240
TCATTGtATT CaCCTCAAAA AGAGAATAGA ATGAAATATT CTACTTATCC ATTGATATAT	300
CTAACGATTG ACTTACAAA TTGAAATAAA AGTCTAAAAG TCAGTTGGTG TGATTATTAT	360
TATAATACAA AAAAGACCTC CATTCTTTTA AGAATAAGGT CTTTTTTTTC GTAGTCGCCT	420
TCTAGGCAAA CTAATTCAGT TAGCCGTATT CTAACTCAA TAAAATGTCA AATACTCATA	480
GGTGTATTTT TATCACAAA TTTACCACAG ATTTGCGGTA TTATTTACAT ATCTTAGATT	540
ATTAAAAATG ACAAACAGC TATTTAATGC GTATTGTATG ATTTTTTTGT TCCTTAAATT	600
GTTATGGAAT TTCAGATGGG GAGTGGCAAA TATACGAAGA AAATTAATA TTTATAGCGG	660
TTTTATTTTA TTAGAGAGTC TCGTTAGTTT TAGTATTGTA TGTGTGATAG CAGGGATATT	720
CTCCCTGACT ATTACACAAT TAATTCAGCA AAATTATCAA CGGGAGCAAG AATTAACGCG	780
CACTCGTCTA GGATACGAAG CAATTTTATT TTTAGAGCAG ACGGGTGATT TACGTTTCAA	840
AGAGCGGATA TATCAAGGAG AACTTACCG ATTTTCTTTG ATGGAGAACG ATGGAAGAAG	900
GATACTAAAA GTCACAGATT CAAGAGGAGC GATTTTGATT GGTCAATAAA AAATATGCGG	960

GTTTTACGAT	GTTAGAGTGT	TTAGTAGCAC	TAGTCGTTTT	AAGTTGTATG	TGTCAATTGT	1020
TTCAGTTGAT	GATTCAACAA	AGTTTTATCG	GAAATCAATA	TTTGAAAAAC	aATGATTCTGA	1080
AATCCTGGCa	TATTTTTTTA	ATTCAATTAG	AAAAAGAATG	TCAAAAGCTC	GTTTTCCAAA	1140
CAGGCTCTGC	TCcAAGAGAT	TTCTTTCCTC	GACTCAGAAA	CAAACAAGAC	TATTTCAATT	1200
CAAATCaAG	AAGATAAAAT	TATTAAGCGA	GTCAATGGTA	AAGGCTATCA	ACCGCTGTTA	1260
ATTGGTATCm	AAAATGGACA	GTTTAAAAAT	GAAGGACAAT	CGTTTACTTT	GGAAGTTACT	1320
TTTACTTCTG	GAAAAACATT	TGATAGTTTT	TTTCCGATAA	AGGGAGAATT	TTAATGAGGC	1380
AAAAATATTC	AGGAAACTTA	TTGTTACACG	CCATGGCCAT	TGTTTATTTG	ATGAGTTTTTC	1440
TCGCCCTTCA	GTTACTAGAA	GAACGTCAGT	TAACACAAAA	ATTTACGCAA	GCTACCCAGG	1500
AATACTATGC	AGGGAAAAGT	ATCTTTCATT	TATTTCTTGC	AGATGTTAAA	CAAAATAGAC	1560
GAAAGTTAAA	AACAGAAGAA	AGGCTCGTAT	ACGCGCAAGT	GACCCTCGAT	TATACATACA	1620
AAAATGAACA	ATTAAGAATA	ACTGTTTTAT	TAAACAAATC	TGGTCGAAAA	TACCAATATC	1680
AAGAGAGAGT	TTCTCATCAA	AAAAAAGCGG	AAACAATACT	GGAATAGCGT	TGTTTTTCTA	1740
CAAATTATTA	GATTTTAGAA	GAAAGACAGT	AAAGGCATTG	ATTATTCTTT	CTTTTTTcGT	1800
TAAAATAAAA	CAGGTATGGA	AGTTGTTACA	TTTGAATAG	AATAGAGGTG	CTAATTTGTT	1860
CCCAGAAAAA	aTTGAAGAGG	CTTTCGAGCC	AATGGAACAA	GCAATCCAGT	TATTGCAACA	1920
GTCATTGGAT	ACTTCTTTTT	TAGATGCGTA	TATTGAAAAT	GGTGAAAATA	TTTAGATGA	1980
TTTTCAAGTA	CGAGTACTTG	ACGGAGTACC	TAATCCAGAG	ACAGTGAAAC	AATTAGAAAC	2040
GCTGTATCAT	ACGATTAAAA	AAATTGATTT	AGCTCCTGAA	GATGTGCGAC	GCTTGTCCCA	2100
ATTACTTTTG	CTAAAAGGAA	CGAGAAAAGA	GCAGCTTCAA	GCAAATCATC	AATTGACGCC	2160
AGACGGTATT	GGCTTTTTGT	TTGTTTATTT	AGTAGAACAA	TTAACAAATA	AATCAGAGCC	2220
ATTAAAAATT	TTAGATCCTG	CTTCAGGAAT	GGGAAATTTA	CTGTTAACGG	TTCTGCTGAA	2280
CCTTGAAACA	GCTGGTTACA	AAGTTTCTGG	TTATGGTGTG	GATATCGATG	AAACACTTTT	2340
GGCAGTATCT	TCTGTGAACA	ATGCTTGGTC	ACAAGCCAAT	ATTCAGTTGT	TCCATCAAGA	2400
CGGATTGCAA	GATTTGTTAT	TGGATCCAGT	TGATTTAGCA	TTAAGTGACT	TGCCGATTGG	2460
CTATTATCCA	AATGATGAGC	GAGCAAAGGG	ATTTGCTGCT	GCTGCGGAGG	AAGGCCATAG	2520
TTATGCGCAT	CATTTATTAA	TGGAACAAGC	CATGAAATAC	GTGAAGCCAG	CAGGGTTTGG	2580
CTTATTCTTA	ATCCCGACCA	ATATTTTAGA	GA CTGAGCAA	AGTGAATTCT	TTAAAAATTG	2640
ATTGACAAAA	AATGTCTATT	TACAAGGAAT	GATACAGTTA	CCTGATGAAT	TGTTTAAGTC	2700
AGAACAATCA	CGTAAAAGTA	TCTTGCTTGT	TCAAAACAAG	GGTGCAGATG	CTGAGCAAGT	2760
GAAGGAAGTT	TTGTTAGCAA	AACTTGCCCTC	TTTGAAAGAT	ATCAATAAAG	TAACGGAGTT	2820
CTTTAAACAA	TTTGAAGCTT	GGAAAGCTTC	AAATTTAAAA	TAAAATAGTA	ACGAAGAGGA	2880
GAAAATTATG	TCTAAAACAA	TTGCAATTAA	TGCAGGTAGT	TCAAGTTTAA	AATGGCAATT	2940

615

ATACCAAATG CCAAATGAAG AAGTGATCGC TAAAGGAATC GTTGAACGTA TTGGCTTAAA	3000
AGATTCTATC TTCACAATTA AATATGGTGA AGGACAAAAA TATGAAGTAA TTGTTGATAT	3060
TGATAACCAT GAAGTGGCAG TAAAAATGTT ACTAGACCAA TTAATCGACT TAAACATTTT	3120
AGGTTCTTAT GATGAAATTA CTGGTGTGG TCACCGTGTA GTTGCTGGTG GAGAAGAATT	3180
TAAAGATTCT GTTGTATTAT CAGACGAAGT ATTAGAAAAA ATTGAAGCTT TATCTGAATT	3240
AGCACCTTTA CACAACCCAG CGAATGCGAT GGGGATTAAA GCATTTAAAC ATATCTTACC	3300
AGAAATCATT AGCGTGGCAG TCTTTGATAC GTCGTTCCAC ACAACAATGC CAGAACACAA	3360
CTATTTATAT AGTGTACCGC GTGAATACTA TGAAAAATTT GCTGCACGTA AATATGGGGC	3420
ACATGGAACA AGTCATCGTT ATGTCTCTCA ACGTGCTGCA GAAATGTTAG GTCGTCCAAT	3480
TGAAGACTTG AAAATTATCA CTTGTCACCT AGGAAATGGT GCTTCAATCA CTGCAGTTGA	3540
TGGTGGTAAA TCAGTCGATA CATCAATGGG CTTCACTCCT CTTGCTGGCG TTACAATGGG	3600
GACTCGTTCA GGGGATGTTG ATGCGTCACT ATTACCTTAT TTAATGACAA AACTTGGCTT	3660
GACTGATGTG CAAGACATGG TTGATATTTT AAACAAAAA TCTGGTTTAT TAGGTTTAAC	3720
AAATGGTTTA TCAAGCGACA TGCGCGATAT TCAATCAAT CTAGATAAAC CTGAAGTTCA	3780
AACAGCATAC AACATTTTTA TCGATCGTAT TCGTAAATAC ATTGGTAGCT ATGTCACGTG	3840
TATGAATGGC GTTGATGCCA TTGTCTTTAC TGCTGGAATC GGTGAAAATG CAGTTGGCGT	3900
TCGTAAAGAT ATTATTGATG GCATGACTTG GTTTGGTTGT GAAATCGATG ATGACAAAAA	3960
TAATGTACAC GGTGAAGAAG CAGTTATCTC AACTGATGAT TCAAAAATCA AGCTTCTATT	4020
AGTACCAACT GATGAAGAGT TAATGATTGC CCGTGACGTA GAACGTCTAA AATAATTAA	4080
CGAACACACC TTGTGAGATG CTCTCACAAG GTnTCTTAAT AGGTAAATAA GAAGAACTG	4140
AGCCAATAT CAAAAGTGAT AGTTGGCTCA nTTnCCTT	4178

## (2) INFORMATION FOR SEQ ID NO: 94:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21252 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

CAAAATCAGC CACCTATGGC CAGTCTATCC TGCCAAAGCG TTTTATTTC TTCATCGGTT	60
GCTTTTCTw GTTCACGATA kTTtCTCaTT TyTTTGCTGA ATAATTCACC GTTACTAGGG	120
GGCGTATATG TAATAGCATT CGCTCCGGCG ATAATGGTTT CATGAATACT CGATTCAGTT	180
GGTCCTCCTG TTGCAATAAT CGGTAATTCT GGATACTTGC CACGGAAATA GCGTACGGTG	240
GCTGCAGTGT CTTTTCCTGC ACTGATATTA ATCATATCCA CGCCAGCCGC AAGTTTTCa	300
TCAATTGGTG TATGaATAGA AGTTAcTGTG CTGATAATAG GAATATCGAT TGCTTCATCC	360
ACCATTTTAA TTGTTTCAAT TGGTGTAGGA CCATTCAAAA CGACACCCAT TGAGCCTTGT	420

GCTTCTGCAA	ATAAACTCAT	ATAAGCAGAA	CGCATGCCTT	GCGTTAAACC	GCCGCCGACT	480
CCTGAAAAGA	CAGGAATATC	CGCTGCTTCA	ATAATGCCTT	TTGTAATCGC	TGGATGCGGT	540
GTAAAAGGAT	ATACAGCAAT	CACTGCCTGT	GCATCCGTAT	TGCGGATAAT	CGCAATATCT	600
GTCGTAAAAA	TAATTGATTT	TATTTTACGG	CCAAAAATTT	TAATGCCACT	GGCTTGCCGA	660
ATAATCTCTG	GCATTTTAC	AATATCTTGT	TTCAGCGGCG	TCATAATTC	TGGAATCCAC	720
TTTTCGCTCA	TTGTCATCAT	TCCTCCATGC	TTAACGGTCG	TATTTATTTA	AATTGTATTC	780
TTCAATTACA	TGAATGATCT	TATCAGCGTA	TGTAGGATCC	GTTGCGTATC	CAGCATCTTG	840
AAGCGCTTGT	GCAGCCTCTT	TATAATTTTT	GGCTAACAAAG	ACTTTTTCAT	ATAAATTTGG	900
ATCCCAATCG	ACCCCATCAA	CAAATAACTT	GGTATGATCT	CGCATGGATT	CTTCCCAATT	960
GTCATAAACT	CGAAAATCAC	CCTGAATCGT	AATCCAGACC	TCATTGACAT	ACTCTTTTGT	1020
TTCTAAATTG	AsTTTGGTT	GATTCCCATA	TGCTTTAATC	CCAAAAAGAT	TGTTATATTT	1080
CGATGCTAAT	TGACTkTGTC	CAAAATTTGA	TTCTAAAACT	GCTTGCCCCA	TAATAATACT	1140
TGGTAATACC	CCGTAAGATT	CTTGTAATTC	ACGTGCATaC	GGTGTGaTTT	CAGCAAcAAA	1200
TTCTTCTTTT	GAGATGTTTT	CATTTTGGCG	ATTGACTGAC	TCGCTTTGTC	CCTTATCTGA	1260
ATCaGAAAGG	ACAGATAACG	AAAAGACAAA	GGCCAAACCG	ACGACAATA	CGCCTAAGAA	1320
TAACCTTTGGC	AAATTCAGCT	GCTTTTTTCT	TTTTTTCTGA	TAaGTTGgCT	CcACTTCATT	1380
AATTCCCCaC	TTCaTCTTTT	aTTTTCCgAt	AAAATGTTTc	aCTAtCTTTc	ACTTTcgtGA	1440
tTGatTTCaT	TTGtTaCTAt	aACCGAATAA	TTGTATCATT	TTACAGAAA	ATTATCAACT	1500
TTCACTACTG	CTTACGCTCG	TGATTTCAGT	ACTTGATTCA	GTGGTAGCAG	TTGAAGAAAC	1560
TGTACTACTT	TCAGAAATTT	CAGTGGATGA	TTTCACTGTT	GTTTCAGTTG	ATGTGTGAAC	1620
AGATTCTTCA	ACTGGCTTTT	TGCCAAAAAA	GACAAAAATT	ACCGCACCAA	ACATAATAAT	1680
CACAGATAAA	ATCCCAGCAA	TTGTGTATTT	CAAAAGACCA	TCCTTCCTAT	TCGTTATTCA	1740
CCTACGTAAG	TATCCATCAG	ATTTTCATTG	ACCCAATCTA	ATAATCAAC	TTTTGAACCT	1800
TCCAATTGTC	GTTGAATCGT	AGCTGGTAAC	CGGAATGTAA	TAATATCATC	AAAACCCCAT	1860
TCACCATAGG	CCAGCGTCAC	TTTCAAATCA	ATATAACCAG	CAGCAATTC	TTGTTCTTGT	1920
TTATTGTTAG	GAATTAATTC	TACAACGCC	AAATCAGATT	GTAGCCATTT	TTTGGCAATT	1980
GTTGTTTTTA	ACCGTTTATA	TTCCGAAACA	GCTAGGTTAC	GTTTTTCCGG	AAATAAGATG	2040
GTTTGGCGGA	GTATTTTTTG	AGTCAGCATC	CACTCATAAT	CCGTAAACAA	ATTCTTAATT	2100
TTTTGCATGA	CTTCATTGTC	CAAACCAAT	TGGCCTTCTT	TCCAAGCTTG	CCAATCGGCT	2160
GCTTGAATTT	GATAAGCGGA	GTAAAAATTT	TCTTCTGTTT	GATACTGTTC	AATAATTGCG	2220
TCAACAATAA	TATCTACATA	AACCTGaTGC	ATAATCATTC	GCCTCCATTT	TCTATTTTAC	2280
TGAAATTCTC	TAAAAAAGC	ATCTTTTAT	CTCTGGaTTT	TATCATATTT	ACAAAATTGT	2340
CCTTGTTTTT	TAATTTCTCT	TGTTTACAGT	ATAATAAATA	ATGAAGAAGA	TTTGAAGAAA	2400



TTGCATAGAT	TAGGAGGACT	TCTTTTGACA	AAAAAATTAT	ATCAATGGGG	CATCGTAGGA	2460
TTAGGTGGAA	TTGCCcACGA	ATTTGCTTCA	ACGTTTAAAC	AAGAAACTAs	CCAaCTTGCA	2520
GCTGTTGCTT	CACGGACATT	GTCTAAAGCA	GAACAATTTG	CCGCTGATTA	TTCAATCCCT	2580
AAAGCCTATG	GTTCTTATGA	AgAAtTGCTA	gCaGATGaAA	CAATCGATGT	TGTCTATATA	2640
GCAGTGCCAA	ATAACAACA	TGCTGAACAC	ATTTTAAAG	CGCTCCAGGC	CGGAAAACAT	2700
GTGTTGTGTG	AAAAAGCTAT	TACGATGAAC	GCGGCAGAAT	TAAAAACAAGC	CTTGGCAGTT	2760
GCAGAAAAAA	ATGACCTTAT	TTTAGCAGAA	GCCATGACTA	TTTTTAACAT	GCCACTATAC	2820
GACGAATTAC	GTCGCTTGAT	GGATGCTGGC	AAATTTGGTC	AGTTAAAAAT	GATTCAAGCA	2880
CCATTTGGAA	GTTACAAAGA	ACCCGATCCT	AAAAATCGCT	TTTTTAATCC	TGACTTAGCA	2940
GGCGGCGCA	TGTTAGATAT	TGGTACTTAT	GCGGTTTCGT	TTGCTCGTTG	GTTTCTCAGC	3000
TCACAGCCAG	ACGTTTTAGC	TTCTAGCATG	GTTCCCTTCT	CTACTGGCGT	TGATGAACAA	3060
TCGGTGACAT	TGCTTCGGAA	TAAAGAAAAT	GAGTTGGCTA	CTATTTCAAT	AACATTCCAA	3120
GCAAAAATGC	CAAAGCAAGG	AATTGTTGCT	TTGAAGATG	CCTATATTAC	CATTACAGAT	3180
TATCCACGCG	CAGATGAAGC	GATTGTTTAC	TACAACGACG	GTACGGTTGA	GACCATCGTT	3240
TCTGGGTACT	CAGCTGAAGC	CATGAATTAC	GAAATCTACA	ATATGGTTGA	AGCAATCAAA	3300
GGTACACAAC	CCAATCGCTC	CTTATTTTTT	ACGACAGAAG	TCATTGATAT	TTTGGACCAG	3360
ATGCAAAAAT	TATGGCAGAA	TTAAAAAAA	GAAGTTCCGC	TTTGCTGAA	GTTCTTTTTT	3420
TTATTTGTTC	TCTATGGATA	AATACGGTTT	AACATACGTG	GGAATGGAAT	CGCTTCACGA	3480
ATGTGTTTAT	TACCAGTAAT	GAACGTAACA	GCACGTTCTA	AACCTAATCC	GAAGCCAGAA	3540
TGTGGCACAG	AACCGTATTT	ACGTAAATCT	AAATACCAGC	TGTAATCTTC	TTCTGATAAG	3600
CCAACTCAG	CTACTTTTTT	TTTCAAATAA	TCATAATCAG	TTGCACGCTC	TGAACCAACG	3660
ATAATTTTAC	CATAACCTTC	TGGTGCAATT	AAATCGGCAC	AAATTACAAC	ATCTTCACGT	3720
GTCGGATGTG	GTTTCATATA	GAAGGCTTTA	ATTGCTTTTG	GATAATTTAA	AATGAATGTT	3780
GGAACGCCAT	AGTAATTTGA	AATAAAGGTT	TCATGTGGGC	TACCAAAGTC	TTGCCCCAT	3840
TCAATGTGAT	CATAATCATT	GTTTGCTTCT	TCTTTTTGTA	ATAACTCAAT	CGCATCATCG	3900
TAAGTAATAC	GTTTGAATGG	CTCCGAAACA	TATTTTTTCCA	GTAACGCTGT	ATCACGTTCT	3960
AAAACATCCA	ATGGATATTG	TTGATTTTTT	AATACCGATT	TAATTAAATG	TTTTACGTAG	4020
GCTTCTTGAA	TATCTAGTGA	TTCTTCATGA	GTTGTGTATG	CCATTTCTGG	TTCAATCATC	4080
CAGAATCCG	TTAAGTGACG	ACGAGTTTTG	GATTTTTCCG	CACGGAAAGT	TGGTCCAAAA	4140
GTAAGACTT	TTCCAAATGC	CATCGCACCA	GCTTCTGCAT	ATAATTGACC	TGTTTGTGAT	4200
AAGAATGCTG	GTTGTCCGAA	GTAATCTGTT	TCAAATAATT	CTGTTGTGCC	TTCTGGTGCT	4260
GAGCCTGTTA	AAATTGGGCT	ATCAATTTTA	ATGAATCCTT	GTTGTTTAAA	GAACCTGTAA	4320
GTAGCACGAA	TAATTTGCTT	ACGAACCTAAC	ATGATCGCAT	GTTGTTTGCT	TGAACGTAAC	4380

CATAAATGAC GATGATCCAT TAAGAAATCT GTGCCGTGTT CTTTCGGCGT AATTGGGTAA	4440
TCTTCTGAAG CGCCAACGAC TTCTAAATCA GTGATATCCA TTTCATAGCC GAATTTAGAA	4500
CGGCTATCTT CTTTGATAAC ACCAGTAACC ATGACTGCTG TTTCTTGkGA TAAATGTTTA	4560
ATTTCTTGGA ATTTTTCTGT ACCAGCTTCT TCACCAAAGG CTTCAATGAA ATTTGGTTTG	4620
AACACGACAC CTTGGAAAAA GGCTGTCCCG TCACGTAATT GTAAAAAGGC AATTTTCCG	4680
CTTGAGCGTT TGTTAGCAAT CCAAGCGCCA ATTTTACTG TTTCTCCAAC ATGTTTGT	4740
GAATCAATAA TTTGAATTTG TTCCACGTTT GTCTCTCCGT TCATTTcTTA GTTTATTCTT	4800
TTAGGCGTTG TGCTGAATTC GTTTCTTTTC TATAAAGTCA TTCATTGCT CAATCGCTTT	4860
TTCTAACGTC ATTAAATCTG TCGCATAGCT CATCCGCACA TGTCTGGTG CACCAAAGCC	4920
CTCTCCCGTT ACCAGAGCAA CATGCGCTTCT TTGTAATAAA TCTTCCACCC AGTTAGTCAC	4980
GTTATCATAC CCACAAATTT CTAACGTTTT TTTGACATTC GGAATAGAT AAAAGGCGCC	5040
TTCAGGTTTT ATTAGCGAAA CGCCTGGCAA AGCTGCTACT TTTGGATAGA TATGATTTAG	5100
GCGTTTTTCA AAGGCCTGaC GCATGTCTTC AACTGTGAC TGTTACCTG TTAAAGCCTC	5160
AATTGCCGCA TATTGGCTAA CTGCGACTGG ATTACTGGTG GATTGTGAAG CGAGTTGTGT	5220
CATTGCTTGA ATAATTTTTT CATCCCCGAC TGCAAAACCA ATGCGCCAGC CTGTCATTGC	5280
ATATGtTTTG GAAACACCAT TAATAATGAT TGTCTGTTGA CGAATtGCTT CTGAGATGGT	5340
TGCAATTGGT GTGAACCGAT GTCCATTGTA CACCAAACGA CCATAAATAT CATCTGCAAT	5400
AATGAGGATA TTATGAGCAA CTGCCCATTC ACCAATTTGG CGTAATTCTT CTTCGTATA	5460
AATAACACCA GTTGGGTTAG ATGGTGAATT TAAAATGATT GCTTTTGTTT TGTCCGTTTCG	5520
TACAGCTTCT AATTGTGCGA CTGACACTTT AAATGATTGC TCTTGCGTGC TTGATACAAA	5580
TACTGGCTTG CCTTCTGCTA ATTTGACTTG TTCACCATAA CTTACCCAAT AAGGGACAGG	5640
GATAATCAcT TCGTCCCCAA CGTTCAAAAT GGCTTGAAAT AATAGATACA GCGCATATTT	5700
GGCACCATCT GTCACAATAA CTTGCTTGGG TTGATAACGA AgTTGATATT CTCGTTCCAC	5760
ATAAGAGACA ATTGCTTGCT TTAATTCAGG AATGCCGCCA GAAGGTGTGT AATAACTTGC	5820
TCTTCCATCC TCAATTGAAG CAATTGCTGC TTTTGAATA TTTTCGGGG TTACAAAATC	5880
TGGTTCCCCA ACCGTTAAGC TTAAACATC TACACCTTTT GCTTTTAATG CTTTCGCCTT	5940
GGCGGcAGCT GcTAATGTAA CAGAAgGAGT CAATTGCTGT GCTCGATTG ATACGTTTCA	6000
ACACCTTTCA CTCCCACTTT CATCAAAACC TATCTAAAtT AAATATCTTT GACAACGTTG	6060
ATCTCTTCGC CTTTATCAAA GCTCATCAGA TAATACGTCA GACTCCTTG ATCATTTTCG	6120
GTGACCACTT CCCATGCAGG GCGCTTATCA TAAAGCCCTA ACGTAGCTTT CATCACCGAT	6180
TCATCGGGAT GGGCTTTAGC AACAAATGTCT GTTATTTGTT GTTCAGTTTG ACCATCTTTT	6240
TGATTTAAGA CCGTGACTTT TTCTCCCGAT TTAGGGACAA TCACGACAAT CTCTGTGCCT	6300
TTATCATTTT TGCCGGTTAA AGTAAAATAC GTATTTTGC GTGTAAACCA GTAAAATTTA	6360

TCAACTGTTT CTAAATTAGC AGAAGTTTTG GCAATTGCAG TTGCTTCTTT TTTGGCTTGG	6420
GcCATTGGTT GATTGGtTCG TACaAAAAAG ATAATTATTG TTACAAGAAT TGCTAAAAGG	6480
ATTGCCAATA GGGTTAAAAG GACAGTAATC CGTTTCTTCT CTTTCTTTTC TTCTAATTCC	6540
TCCAAAAATC CTTCCATCCT CTCTAAACTT ACTGGGTCTT ATTATAGCAG AAAAATAGCG	6600
GCCTTTCACA AATATGACCT AAAAATAATC GATTATTTGT TTA AAAAATTC CTTTAATTCT	6660
TCTAAAATTT CAGGCATGGG CGCTTCTTTT ACTGAGAGTC CCTTGGGAAA CGCTTTGATT	6720
AATCGTTTGC CATATTTTGC AGTGACTAAA CGACGATCCA GAATCAGCAA TGCCCCCTCTG	6780
TCTTCTTCAG AACGGATTAA ACGGCCCAGT GCTTGTGCGA AACGTAAGGC TGCTTTCGGT	6840
ACCGCTTCTT GAGCAAATGG TTGTATCCCT TCTGCTTCTA AATACGCATA TTTTGCTTGC	6900
ACAAGTGGTC GTTTCGGGTT TTCAAAGGC AAACGCGTGA CTATCAATAA CTGTAAAGCT	6960
GTTCTTGGA AGTCCACACC TTCCCAAAAG CTATCGGCTC CTAGCAAAAC ATTTGGTTTT	7020
CCATTAGAGA ACCGTTTAAG AATTTTTTCA CGACTGCCG TAATCCCTTG GGCGAGGACC	7080
TCTTGTCCTT TTtCCAATAA TCTTGGCTGG ACTTGATGAT ACACGGTGGA CAATAATTCA	7140
TGCGAAGTAA AAAGAACCAA CATCGAACGT TCCTCTTGCG CCATCAATTT TTCAATGGTA	7200
TCTGAAATGT AcTGACCCAT tTCATAGGmA GAGGCATTTT GAATGGCTAC TGCTTCTTCC	7260
GGTACATAGA CTCGCGCATT TTCTGCATAG TTATAAGGCG TTGGCAGGGT TTTGAACGTC	7320
ACCTCTGTTA AACCTAACCG TTGCGGTAAA TACTTTTTAT TGTTGCCAAA CTTCAATGTG	7380
CCGCTTGAT AAATAATTCTG TTCATACCGC GGATACCACT TTGTTTCAGG AAGCAAAC TA	7440
GCTTCTAAGT CATGAATTTC AATCAAGCCA TGGTCTTCAT TTTGTGGTAA ATATTCTTTT	7500
ACCCAACGTG GACGCCAATC ATCAAGATAA ATGGTAAAC AGTGATGAAT ACGTTCAATT	7560
TCATCAAAAA ATTGAAACAA GCGAACAAAA ACAATGCGAT CAGCCAAACT AAAGGTTTCC	7620
AACTGACTCA TAATTGTTG CTGTAACCTA CTTTGAATAG TTAACAGCTC CTGCAGTAAA	7680
AGCTGGAGCG TTTGACTATG TTTTCTCCC GCAGCAGATA AATAATCAAT ACGTTCTTTG	7740
GTTAAAAGCA ACTCCTGATG GTCTATCCGC TTCTGTGTTT TACTAAATAA ACGATGCAAT	7800
TCAAAAAGAA AATCTTGCCA TTCATCCACA AAATCCAGTA AACCTTGCTG ATAAATTTTT	7860
ACAAACCGCT GCTCTTCAGA AAAGTGTGA AAAATTTGCT GTAAAGCAAC AAAAAGTTGT	7920
TCTTTTCAA TAAATGCGT GACTTGTTTT TTAGTATCA CACTGTTTAC TCGTTCGCTC	7980
GCCATCCGAC TAGCGATGTC TGGTAAATGA TGTGCTTCAT CAATAATTAG CCACGGACTT	8040
TCAGGTAATA ACGTTGTTT TCTTAGCGAC TCTTGCGCCA AAAATGCATG ATTAACAATC	8100
AACACGTTAC TTTGAGCGAC TTTTTTATAC AGATGTCTTA AAAAGTCAAC TTCATAAAGT	8160
TCATCTTTTT TTGAAAGAAA ATCCAATCCG CGATGTGTAA CTTCTTTCCA GAATAAATGA	8220
TTTAAATTCG TCAATTGTAG TTCATCTAAA TCGCCCGTTT CTGTTTCTAG CAACCAGACG	8280
AGAACGCCCA TTTGATAAAG AGCGTATTGT TTTTGCGGAG TTGGCTGATT CAACGTTGCC	8340

TTAAAGCATT GTAAATCAAG ATAGTGTCGA TGGCTTTTAA TAACAATGCC TCGTAATTTT	8400
CCTTGGCAGA TTTGATTGGC TAATGGTAAA TCTTTTCCA CTAATTGATT TTGTAAAACA	8460
ATCGAACTG TTGAAATGAT GACTGGTTTT tCTGGCGTTG CTAAATAGCT CATCGGCAAC	8520
AGATAGCCTA ATGTTTTTCC CGTTCCAGTC GCCGCTTCAA TAAATAAATC TTTTGTCTGT	8580
CCTTCTGTAA AATGATCATA GACTAAATTC ATCATCCGAC TCTGTCCGC ACGATAAGCG	8640
AAACGTTTCG CAAACAGTTT TTCTTTGGCT TTTTGTAGTT TGGGATACGT GCTAGTCTGA	8700
TAGAACGTTT CTTCAAAAAG AGGAACCTCT TTTTGTGCGTA AAGCAATGCC ACTGACCACT	8760
TGATAAGCTG GATTTAAATGG TGTGACTTGT TTTTTCATTT GTTCATACGT TTGTGAATA	8820
AAGGCACTGG TTTCCCGCGC TGTCTGTTGA CTTAATTCCG CAATTTTTTC CATAGTGACT	8880
AATGGTAAG ATTTTCATTTT TTCTTGAATT AATAACAAAA GTTCgGCTGT AACCTGCGCA	8940
TCACTATCTG CTTGATGGGG ATTCTCATGA GACAAACCTA AGCTTTCCGA CAAATCACTT	9000
AAACGGAAAC TTTTTTCTGT CGGTAAAAAA ATCTGTGCCA ATTCCACTGT ATCAATGGCT	9060
GGAATAGTAA GTGGCGGCGT GCCACACCGT ACCAGTTCAC GAGCTAAAAA ATTATAATCA	9120
AAGTGGACGT TATGCGCCAC AAAAATAGTA TCTCCAGAT AATGATAAAT TGTATGTGCT	9180
ACATCTTCAA AATAAGGCGC TTTTGAAC TGCCTATTAC TAATCCCCGT TAACTTTGA	9240
ATTTGTTTTG GTACTACTTG ATTGGGATTC ACATCTGTCTG CAAAGTTCGC AATAATCTTT	9300
CCATCTTGAA CGAGAACGCA TCCAAACTGA ATGATCCGAT CACTTGATAGG ATCTGTTCCA	9360
GTCGTTTCAA GGTCAACTAC TGCATACACT TGTTTGTCTT TCAATGACTT CACGCCTTTC	9420
ATTCATCTTT GAAAATTATA CTACAATTTA TAGGATTTTC ACTAGTATTT CAAAAATTCT	9480
CTTGAAATCT TTTCCAGAGA TAATAAGAAA GcGCCTGAAA CAAAAATCCA AAGTGATTTT	9540
TGTTTCAGGC yTAAAAGCTG ATAAATTGAA CGGTTTTGTC TTACCTTCTC CTAAATAAAC	9600
GATTAATAAC AGTCAAGGAT TTTCTGCATA TACTGGTTTA ATTGCCATA GCCTTCCAAT	9660
TCTCTGATTA CTGGGAGCGG ACAACGAACC TTTTATTAA ATAGTGTCG CTTTCAACA	9720
TATTCTGTAA tTCGTCTAAA AATTTATCTA ATCCTTTTTT TAATTTTATA CAACTTAACT	9780
GATTGCAAAT TCGCACGTCT CCTGCCTTTT TCTTTTGAAC TTGTTTAAAT GATTGTGAGA	9840
ATGCTTTTGC TTCAAACATT TGTGCTTCTA ACTGTCTTTT TTCAATAGTT ATTTTCGGTA	9900
CTGCTTGATT TCCATCAGTA TCCTCTAAAT GCGGACTCAT AAAGCGACGA TAAATTCCT	9960
CTTCTAATTC GTCCAATTGT CGTTGAAAGT CCGCAGGAAT TGTCCGACTT GATGAATTTT	10020
CTTTCTCACT TTCCATAGAG ACAGCTTCTT TTTCTGGCTT CTGCTCCTCT TCTTTTTCGG	10080
CAGACTTATT CACTGAATTT AGCGAGCTTT TTTCTGTTTC TTGTGGAAAA GTAAATAAaG	10140
GTTTTGCCGC GGTACTCTGC TGGTTTTCTT GTTTTTTTTC TGGCAGCGCT TCTTTAATTG	10200
GCGCTGGTTC ACTTGTGCT ACAGTTGTTT CACCTTCTTG ATTTTCCCGA TTATCAGGCG	10260
AGAGATTGAC GTCTGGTTGT TCCAGATTAG TATCCTCTGA TACTTTCAGC GATGCCTGTT	10320

CTGCGGCTAT CACAGGAGAA GCTCTTTTCA AATGGATCGT TTTTATTTGT CTAGAAATGA	10380
CGGGATTCTT CGTTTGATTA ACGGCTTCAT TTAAGGAGGC TAAATTTTCT TTTAAAAGTG	10440
TCAACATTTC ACGTAAGACA GCAATTTCAC CAGCTAATTC CGTTAATTGA TTTTCTTGCG	10500
GCACACTTGG CACATTGGTT TCATTCATAC GGACTGTTTT TTTATTTTGA GTTGTGCGTA	10560
TTTCATACAT CCATTGTTTC AGTGCGGATC GTAAGTCACG AGTTGTGCGC TCAAAcTTCC	10620
AAAATAAACA GTGTAGTACA TTAcTCCGTT TCAACATAGA CATAAATCTT TTTAACAATT	10680
AACTCTTCAA ACTGTTCTGA TTCATCCCAA CTTTCGCATA AAGCCTCAA AAATTGTTCC	10740
CCTTCTGCTA AATTTTGATA CTCTTTCGCT AAATAAAGCT TGCGrATTAA GTAAAAaTTA	10800
TACGGTGTTC CAACATTATC TGCATAAATT TCTCGTCCTA AAAAATCTGG ATTCGTTTCT	10860
TTACAAAGAA AGACCGTTTT CTCTTCTGGT TGTGTTTCAA GTGTTTTAGT CATGAGATTC	10920
ACTTCCTTTT TCTCCGTAA TAACCATTTT TTAATGATAG ATACCTTTAG TATACAGGAG	10980
AAATCTTCTT GATTTTCTCA TATTTTAAAG AGACAATAGA AATTTTAGAG AAAAATCGAC	11040
AGAAAAAAA CGTGAGATAA ACAGTCCTAT TGACCATTTA TCTCACGTAA AATACTTAAA	11100
CTGACTATTT TATCCTTGGA CATGTAACCA ATTATTCTGC TGAATCAATT GGCAAAATCG	11160
TGTCTTTTTT CTCGCGACGT GCAATTGCTG GTAAAATCAT TCCTAATAGA ATTAAGACGA	11220
TTGGTGTAC AATGTTAGAA GCTAATTGGA ACAACCAGCC TTGAGGGTCT GAGGCATACT	11280
CAAATTTAGG CACCATTCCCT AGGATACAAG CAAAGGCTGT AAATAGGAAA CACCAAAGAC	11340
CAAACACAAA ACCAATTTTA GGATTTTAA CAAATTTATA TTCTGAATTA AATTGTTTGT	11400
ATGCTTTGTT TAGAAGCATA TAGGCAAAGA AGACCCACAA GTAACGCATT GGCATAACTA	11460
CAGAATTTAA GTTAGTTAGC CATTTAACCA CTTTCGTTTAT GTCCTTCAAC CCAAACATTG	11520
GTAAAGCGAT AATAATACTT ACTAAAATAC CTGTTAATAA GTAACCATTA ATTAACGTAC	11580
CTTTTTTCGT ACGTTTACGT AACCAACTTG GAACAAATTC TGGATCTGCT TCAGCTAAAA	11640
GAATTTTCAG TGGAGCATCA ATTGAAAAGG CTAGCGCagc CACTTGACCC ACTGTATTCTG	11700
TCAAAGCATA AACAATCATT AAGAGATTAC CAACACCATA ATATGCGCCT AATTTTGGGA	11760
AGGCACTGTA GGCACCATTT GCCATTAAAT CTTTAGGAAT ATTGTCACTT GAGAATAACA	11820
TGCCCATTGC AAATGAACCT AAAACAGCCG ATAAACCAAC CATTGCCGCT AGGAAATCA	11880
TTCTTTTAGG GAACTCTTTG GCTGGATTTT TAGTTGCATT TACGTAAGGA GAAATTTTTT	11940
CAGCGCCACC TACCGCAAAA ACTAACATCG ATACGGTTGT AAAATAACTA AAGTCAAATT	12000
TAGGAATATA GGTGCTGACT TGTGACATGT TTGGCGTCGC AAATTCAACC GTTGAATTAA	12060
TAAATGGTGC GCCTACCGCT AATAAAATGA ATAAATAGA CATCACAAAC ATGGCGGTTT	12120
CAGCTAAACC ACCAATGACT TTTAAGGTGG TTAACCCTTT TGTTGATAAT AACAAGAAAA	12180
CCAAGAAAAT TATTAATGAA ATAAATGATA CCGTTACGAT TGGCATCGTA TTGACTAGGT	12240
TTCCATTCCC TTGAACTGCC CAACCTAAGG CAATTAAAT TGCTTGTGGT TTTTGTGCCA	12300

AGTAAGGAAT GTGCACTACC CAATAAGTCC ATGCTGCATA ATAGGCTAGA CGTTTAGTAC	12360
TGGTATTTTG CACCCACGAA CTAACCTCCAC CATTACTGTC TTTAAATGTT GAACCTAACT	12420
GTCTTACGAT TAACGCATAG GGAATAAAAT ATAAAACTAA AATCAAAAAGC CATGAAGTGA	12480
CCACTGAAAT CCCTTGTTGT GCATAATTAT TCACAACATT TCCAAGACCC CAGACCATGT	12540
TAAAGGCAAT TAACCCGACG GTGAACCATC GTAATTGCTT TTTGTCCATA TAACCAATTC	12600
CTTTCTTATT AAAAAAATA AAGCTCTTCA TTTATCATAA GCATTTTACT TTTTATGCAC	12660
AAGCTCAACA TAATTAAATA AAGATGGCTT TTTAATAAGT AATTAAAAACG CTTCTAGCAA	12720
TGATTTTCTA ATTTATTTGA TGAATAAAAA AGGCCGGTGA ATAAGTTGCT TCACCAGCCT	12780
TTTTCTAAGA GATTTTCTT TTATAACATA TCAATAATG ATAATTGGTT TTCATCTGGC	12840
AAGTCTTTTA AGACACCATT TTCATTCATG TATTCAATTA ACGTTTTAGA AACTTTGCCT	12900
CGTGTGCGCA AATCTTCTTT CGATAAGAAC GGTGCTCTT CTCGCGCAGC AACAAATGGCT	12960
TTCGCTACGT TGAGGCCTAA ACTTGGCACT GCTCGGAAAG GTGCAATCAA GGTATCCCCT	13020
TCAATCACAA AATTTTCAGC ATCTGATTTG TACAAATCAA TCATGCCAAA TTTATAGCCC	13080
CGTTCCAACA TTTCATTACA TAATTCTAAA ACCGTTAATA AGTTTTCTC TTTAGTGGAG	13140
GCGTCCATCC CTTTGtCCGT AATTTCTTTC ATTCGTGCTT TCACAGCTTC TTTTCCTTGT	13200
GACATTGCCA CTAAGTCGAA GTCATCCGCC CGAACGGAGA AAAACGCACA GTAATATAAG	13260
ATTGGGAAAT AAACTTTAAA GTAAGCGACC CGCAACGCCA TTAATACGTA TGCTGCCGCA	13320
TGGGCTTTTC GGAACATATA TTTAATTTTA GAACAAGAAT CAATATACCA GTCAGGGACG	13380
TTGTTTTCTT TCATCGCATT TAAATACGTA TCCCGTAGTT CATCAGGAAT CTTATTCCAT	13440
TGTCCTTTAC GTACCGTTTC CATAATCTTA AAGGCCATCC CACTGTCCAA ACCAGCATGA	13500
ATTAAATAAA CcATGATATC ATCCCCGACAA CCAATTACTT CAGCAAGATT GgctTCGCCA	13560
CGACGAATCA ATTCTTCTGC ATTCCCTAGC CACACATCGG TTCCGTGAGA CAGTCCAGAA	13620
ATTTGAAGTA ATTCTGCAAA AGTTGATGGA TGGGTCTGTT CTAACATACC TCGAACGAAA	13680
CGTGTCCCAA ATTCAGGAAT TCCTAAAGTT CCTGTTTTAG AATAAATTTG CTCTGCATCC	13740
ACACCTAGAA CATCAGGTCC TTCAAAAATA CGCATCACTT CTGGATCGTC TGTTGGGATG	13800
GTTTTTGCT CAATCCCAGA TAAATCTTGT AACATCCGAA TCACGGTCGG ATCATCATGT	13860
CCTAGAATAT CAAGTTTTAA AATATTATCG TGAATGGAAT GGAAATCAAA GTGCGTCGTT	13920
TTCCATTCTG CCTCTTGATC ATCAGCAGGG TATTGAATTG GCGTAAAGTC ATAAACATCC	13980
ATGTAATCAG GAATTACAAT AATACCGCCG GGATGCTGCC CAGTCGTCCG TTTAACCCTT	14040
GTTGAACCTT TGGCTAAGCG ATCAATTTCA GCACCACGCA AATGTAAATT ATGATCTCTT	14100
TCATATCCTT TAACAAAGCC ATACGCTGTT TTGTGGGCTA CTGTACCAAT GGTTCCTGCC	14160
CGGTAGACAT ATTCTTCACC AAAAAGAACT TTCGTGTAGT TATGGGCTTC GGCTTGATAA	14220
TCGCCAGAAA AGTTCAAATC GATATCGGGA ACTTTGTCGC CATGGAACCC TAAGAACGTT	14280

TCAAAAGGAA	TATCGTGACC	ATCTTTAAAC	AATCGAGCGC	CACATTTTGG	ACAGGCTTTT	14340
TCAGGCATAT	CAAACCTGA	ACCATAGGAA	CCATCTTCAT	AGAACTCAGA	ATATTGACAC	14400
TCAGGACAAT	AATAATGTGG	CGCTAAAGGA	TTTACTTCAG	TAATCCCTGT	CATTGTCGCA	14460
ACAAAACCTG	AACCGACAGA	GCCCCGTGAA	CCAACTAAAT	AGCCATCTTC	ATTACTTTTA	14520
TGCACCAGTT	TTTGAGAAAT	TAAATAAATG	ACCGAGAAAC	CATTCCCGTT	AATCGAATTC	14580
AGTTCCTTTT	CTAACCGTTT	TTCGACAATT	TCTGGTAAAG	GATCGCCATA	CAATTCGCGC	14640
GCACGATTGT	AGCTTAAATC	AGTGATCTCT	TGTTAGACC	CAGGAATTTT	CGGCGTATAC	14700
AAATCGTCTT	TTACTGGAAT	AACTTCTTCA	CATAAATCAG	CGACTTTGTT	GGGATTTTCT	14760
ACTACAATTC	GCTTAGCGAT	GTCTTGTCCT	AAAAATTGAA	ATTCCGTTAA	CATTTTCATCG	14820
GTGGTTCTGA	AATGAACCTT	TGGTAAACTA	TGGCGATTTA	ACGGATTAGC	GCCGCCATA	14880
GACCCGACTA	AAATTTTCCG	ATAGATGGCA	TCTTCTTCAT	TTAAATAATG	AACATTTCCA	14940
GTGGCGACAA	CCAGCTTACC	AAGTTCGTCA	CCGATTTTCA	CCAGATTACT	AATAATTTCT	15000
TCCAAATCTG	CTTCATTTTT	CACCAATTCT	TGTTCAATTA	ATGGCGCGTA	AACTGGTTTG	15060
GGCATAACTT	CAATATAATC	GTA GAATTTG	GCACGATTTT	TTGCTTCTTC	GACACCTTTT	15120
TGCATCATTG	CTTCAAAAAT	TTCGCCGTTT	GAACATGCGG	AACCGATTAA	AAGTCCTTCT	15180
CTTAACTTCG	ATAATTGCGA	ACGAGGAATC	CGTGGTACTC	GGAAAAATA	ATCTACATTG	15240
GACATGGAAA	TCAATTTAAA	AAGATTTTTT	AATCCTGCTT	GAGTTGTTGC	TAAATCGTT	15300
GCATGGAAAG	GTCGAGCTCG	TTTGTATGAG	TCTCCTTCGC	CAATATGACG	ATTCAAATCA	15360
TCATGGAAAT	GCATATCATG	ATTCTCTTTC	GCTTCTTTTA	AGAAAATCCA	GCACAAATGA	15420
CCTGTTGATT	CCGAGTCATA	GATGGCGCGG	TGATGTTGTT	CTAAGTTGAC	GCCAAATTTT	15480
TTCGATAAGG	TATTAGACG	ATGGCTTTTG	AAATGTGGAT	ACAAGAATCT	GGATAATTCT	15540
AATGTATCGA	TGACGGGATT	AGCCGCTTCA	GGTATTCCAT	GTTTGCCATA	ACTTGTATTT	15600
AAGAACCCCA	TATCAAAAGA	AGCGTTGTGG	GCCACTAAAA	TAGTTCCTTC	TGAAAACCTCT	15660
TTAAACATGC	GTAAAACCTC	TTCTTCGGAT	TTTGAACCGC	GCACCATTTC	ATCTGTAATC	15720
CCAGTTAAAT	TGATCGTTGT	TTGAGATAAT	GGATGGCCAG	GATCAATAAA	TTGCTCAAAA	15780
GTCTCAATGA	CATTTCCTTT	GTGcATTTTG	ACAGCGGCCA	ATTCAATAAT	TGTATCATAA	15840
ACAGCGGAAA	GCCCCGTTGT	TTCCACGTCA	AAGACCACAT	AGGTAGCATC	TGTTAATTCA	15900
ATATGTTCTT	CGTTGTAAGC	AATGGGTACA	CCATCATCAA	CAATATTGGC	TTCCACGCCG	15960
TATAAAATTT	TAACGCCTGC	TTTTTACCT	GCAGCATGGG	CATCAGGAAA	CGCTTGGGCG	16020
CCGCCATGAT	CCGTAATTGC	GATGGCCTTA	TGACCCCAT	TCCCCGcTTG	CGCCACCAAG	16080
TCACCGACTT	TATTGGTCGC	ATCCATTGTA	CTCATGTTAC	TATGCATGTG	TAATTCTACA	16140
CGTTTTTCAC	CTTCTGGCGC	GTAGTCTTTT	CTTGGCGCAT	GGGCGACTTC	CATCAAATCT	16200
TGAGCATTCA	TGACTAAATC	ACGCATGAAG	GTATCTTCTT	GTACACTACC	ACGAACACGA	16260

ATCCA <del>A</del> CTTT	GCGGTTGAAT	CGCATCAAAA	ACTTGTTTCAT	CTTTTTCCCC	ATTCGAAAAT	16320
TTCTTCACCA	CGAAAGAGGA	AGTATAATCA	GTAATTTTCA	AAATTAAAAT	TTTTCGTTTA	16380
GAACGCAATT	CACGCACTTC	TTTATCAAAA	ATAAAGCCTT	CAATCGTTAT	CCGACGTTCT	16440
TCTTCCAGTA	TGTTTCCCAT	TGGCATAATG	GGTTCATCAT	TGGGAATATT	CCGACCTAAA	16500
CGAATTGGAC	CATCAAACGC	AGGCGCTTGT	TGTTGTTTTT	CTTTTTTCAT	TTGTTTCATGT	16560
TTAACAAGCG	ATTGAGCAGC	TTGTTGTTGA	AAGGCTGCGG	CTTGTTCTAA	TTTTTGCTCT	16620
TCAAATTTTT	TCAGCACTTC	TGCAGCTTGC	TGTTGATCCA	TTTTTGGTTC	AATATGAAAT	16680
TTAGGAAACC	CGTAAGAGAA	ATAAAGTCC	TCAATAATG	GTAAATATTG	TTGCTTCAAA	16740
TAAGGAATAA	CTGCTTCATT	ATCAACAGGT	AAGATAATTT	TCCGATCTTC	AAATTGTGGT	16800
GTCTGTGTTT	TTAAGACTTT	CTGAACTAAC	GGCGTATTGC	ATTGACTATT	TAACAAGGCC	16860
AATTGCCAAT	AATCCGTAA	TTGTTCTCT	GTAATGTTG	TCTGATTGT	GGTTATTTGA	16920
ACGGAAACCT	GCGCAATTTG	TTGAAAAGCT	AGTTCCAATT	GTTGCATAAA	TGTTTGATAT	16980
AGCATCACTG	GCATCAATTC	ATCAAAACTT	AAATGAAATT	CCCATAAGCG	ACTTTGTTGG	17040
TGAACAATGA	CCTTGTCAT	CCGGCCATTT	TGAATTAAGG	GATGTTCTTT	CGCTGTGTCA	17100
TCTAACTGAA	TTTGCCTCAT	CAGTTTATCA	AATAAATCGC	GTGCTTTTTT	TGTCACAAAA	17160
TGAACCTCCT	TCTATCATTG	AAAAAAGGCC	CTCAGACAAA	TCTGTCAAGT	GCCTTTTTTC	17220
GGTTTCTCTG	CATTCGCTTC	TCTATTAAGT	ATACTCGTTT	TTTGTA AAAA	AACAATCTTG	17280
TTTCTTTTCT	AAAAAGAGCT	AAGAAGAGCT	TGAAGCAAAA	ATCACTTTAG	ATTTTTGCTT	17340
CAAGC <del>y</del> TAAA	CGCTGATAAG	TGGCGGAAAC	AGTTTATCCT	AATCATTCTG	TCCGCCCTTT	17400
TCATTATTAT	TCCACTTCTG	ATGTTGTGTT	CATCA <del>r</del> AATA	GAAAGTGAC	TTCTAAT <del>t</del> C	17460
TTCTTTGCGT	ACTTCC <del>a</del> ACA	TTTCACCAGT	ACGTTTGATT	T <del>k</del> GACTTCTA	CCACACCATC	17520
GA CTGCTTTT	TTCCCGACTG	TGATACGGAT	TGGACAACCA	ATTAAATCAG	CATCAGCAAA	17580
TTTTACGCCT	GCGCGCTCGT	TGCGGTCATC	AACTAATACT	TCGTAGCCCG	CTTCGGTCAT	17640
CATTGCTTCA	ACTTCTTG TG	ATAGTTTTGT	TTGGTACTCA	TCTTTCACAT	TCATTTGAC	17700
CACGTGTAAA	TCAAATGGCG	CAATTCTCTG	TGGCCAATTA	ATACCACGCT	CGTCTGCATT	17760
TTGTTCAACG	ATTGCTGAAA	GTAAACGACT	AACGCCGATG	CCATAGCAAC	CCATGATCAC	17820
TGATTTTTC	CGACCGTTTT	CATCTAAAAC	TGTTGCGCCC	ATAGCATCAC	TATAACGCGT	17880
TCCTAATTTA	AAGATATGAC	CAATTTTCGAT	TCCTTTAGTA	AAGGCTAAGA	CACCATTGCC	17940
ATCAGGAGAA	GGATCGCCTT	CTTGAACAAA	GCGTAAATCT	TCGTAAC <del>TAA</del>	TTGGTTGGAA	18000
ATCACGATCT	GGATTTACAT	TTGTCAAATG	ATAGCCATCT	TCGTTAGCTC	CAACAATGGC	18060
GTTGGCTAAA	TCTTGAACAG	CTAAATCTGC	ATAAATTTTA	ACATCTTCGG	AAACATTGAC	18120
TGGACCAATT	GAACCAAAGC	CCGCTCCTAA	GACACGGCGA	GCATCTTCTT	CTGTTGCTTC	18180
ATCTAAGAAA	TCAGCGCCTA	AGAAGTTTTT	CAATTTGACA	TCATTGACAT	CATGATCGCC	18240



ACGAACCAAG	ACCATTACTG	GCTCTTCATC	GGCAATAAAT	AGTACTGATT	TAATAATTCTG	18300
TTGTGGCTCA	ACTTCAAAGA	AGTTTGCCAC	TTCGGCAATT	GTACCCACTT	CTGGTGTTGC	18360
AATTTTTTCT	AAATCTAATT	GTGTTTCATG	TGATTTTTTC	GGCGTATAAA	GACTTGTTGC	18420
CATTTCTAAA	TTGGCAGCAT	AATCACTTTC	TGTTGAATAA	CAAATAGTAT	CTTCCCCAAT	18480
TTCAGAAATC	GCCATAAATT	CTTTAGAATC	TTTACCACCC	ATGGCACCAC	CGTCCCCAAT	18540
AATGGCACGG	AATTCTAGTC	CACAGCGTTC	AAAAATGCGT	GAGTAAGCTT	TTTCGTAATC	18600
ACGGTATGAT	TGATCTAAAC	TAGCTTCATC	GGCATGGAAA	GAATAGCCAT	CTTTCATGAT	18660
AAATTCACGG	CCTCTTAATA	AGCCAGAACG	TGAGCGTTTT	TCATCACGAT	ATTTTGTTTG	18720
AATTTGGTAT	AAATCAAAG	GTAATCGTTT	GTAAGAATTA	ATTTCATCAC	GAATTAATTC	18780
TGTGAAGGTT	TCTTCATGCG	TTGGGCCTAA	GATATAGTCA	CGGTCATTGC	GATCTTTTAA	18840
GCGATACAAG	TTAGGTCCAT	AAGTTTCATA	ACGGCCTGAT	TCTTTCCAAA	GTTCTGCGGG	18900
TAATAACGCA	GGCATTAAAC	TTTCAACTGC	ATCAATTTC	TCAAATTCTT	CTCGCATAAT	18960
CGTTTTTAAT	TTTTCTAAAA	CACGATTGGC	TAAAGGTAAA	TAAGAATAGA	TCCCTGCAGC	19020
GACTTGACGA	ATATAGCCTG	CGCGCAACAG	AATTTGATGA	CTTAAGACTT	CTGCATCATT	19080
TGGTACTTCT	CTTAACGTTG	GAATTAACAT	TTTTGACTGT	TTCATTAAAA	CTCTCCTCTA	19140
TTTTAAATC	AACTATTTTT	GATTTCTTTA	CACACTTTTT	TACTTAAACT	TTCTCCTTAA	19200
AAGAAAAAGC	GTTGAATATC	GTTCCAAGTA	ACTaACACCA	TTaACACCAT	GACAAACCCA	19260
AAGCCAATTA	ACGTAATAT	GCCTTCTTTT	TCAGGACTAA	TTGGTTTTCC	ACGTACACCT	19320
TCAATAATGT	TTAAGACAAT	TTTCCCGCCA	TCTAAAGCTG	GGATCGGCAA	CAAATTAATA	19380
ATCCCTAAGT	TCATTGACAA	CATGGCCATT	AAGAATACAA	CTGTACTTAC	TCCAGCATTG	19440
GATGCTTCTT	CCGATAATTT	AAACATCATG	ACTGGCCAC	CTAGTTTGTT	TAAACTAAAG	19500
CcTGTGAATA	GTGAGCCGAG	TGCTTTAAAA	ATCTGTGTCG	TACTATTCAA	AGTATCCTGA	19560
ATACCGCCCA	TCAATTTTGA	CGGTAAATCG	GTTTTCATAT	AAGGATAAAC	GCCGACTTTA	19620
CCAATCGTTT	GTTTTTCCAC	TTTTTGTTTT	TCTGGTGTCA	CTGTTAGTTG	CTCTTCTTTG	19680
CCGTTACGCT	CAACTACGAA	CGTTAACGGC	TTTTCGGGGT	TCTTCTGCAC	AATGGTTGTA	19740
AAATCTTCGT	ATTTTTTGAT	TTTTTGATTA	TTAATCGATA	AGACTTTATC	GTTTTCTTTC	19800
AACCCAGCTT	CTGCGGCTGG	GCCATTAGGA	ATCACTTGTC	CAATTTGATT	CGTGTTTAAA	19860
TCAGTGACGC	CTCCTTGTA	AAAGACCGCT	AACGTAAACA	GAATAAACCC	TAAGATAAAG	19920
TTATTCATGG	GTCCCGCAAA	GTTCTGCAAA	ATGCGTTGCG	ATAATTTGCG	TGATTGAAAT	19980
TGAACGTCAA	GTGGCGCAAT	CCGCACCTCG	GTTCCATCAC	TTTCAATAAT	CGTTGCATCA	20040
TGGTCAACTT	TATAAACGGT	TTCTTCTTCC	TCGTTTCCAT	TGACATAGCC	CTTGATGAAT	20100
AATTCCTTTT	CAAGATCAAA	ATCAATGACT	TCCATCGGAA	TACTATGAGG	TAATTGTACT	20160
TTTTTACTTG	TATTAATTTT	AACCACATTA	CCCACGGCAT	TTAACTCAAC	AGATAGAGGC	20220

ATACCTGGTG TGATTTCTGT CATGTCTTCG CCCATCCCAG CCATTTCGCAC ATAGCCACCA	20280
ATTGGCAATA AGCGAATCGT ATAAGTGGTG CCATCTTTTC CACGATGCGC AAAAATCTTT	20340
GGTCCCATTC CGATTGCAAA CTCACGAACT AAAATACCCG CTCGTTTTGC AAAATAAAAG	20400
TGGCCAAATT CATGTACGAG GACAAGAATA CCGAAGACAA TAATGAATGT GATAATTGTT	20460
TTCATAGATG TCCTTCTTCC TTCATTCACT GACGTTCACT GCTTTTTATT ATTACATGGG	20520
AAAGTCTCGT AAAACTTGCT CTGTTCCCTT CCCGATAAAC CAGTGCAACT GTTTACTGAG	20580
AACAGTCGCC ACTGTTAGTG AACAAATCTT TCCTTATGTA AGACAAGTTT TATTTTAGCA	20640
TATTCACTCA TAAATAACCA TTGGACTTGA CCACTTCGCC TAGATATTTT GTGTTTTTTT	20700
AGAAAACTGT ATAAATGAGA GGmaTTGATT ACTACGACCA GTATTTATTT TAAAGACCTT	20760
CAACAATAGT AGCTAACTTT AGACCATAAA CTTTTCCGAC AGTTTGTCGT ATTTTATCAA	20820
CACGCTGATT GACATGTGCT ACCATGAAAG TATTAATAGT AATTTTAGGG GGAATGAATC	20880
AATGTATGGA GAATTACATG GTAAAGTAGC CGTTGTCACT GGTGCAGCTA CGGGTTTAGG	20940
ATTATTTATT AcTCTGCGTT ATATCTTAGA GGGCATGAAT GTTGTGCTG ACTATGTCGG	21000
TGAGCTGCCT AAAGAATTTG AAGATGTCCA AGCCAAGCAC GCAGATCGCG TAAAATTCGT	21060
GAAGGCTGAT GTCTCTAATG AAGAAGATAT TAAAGCGTTA gCTGAAACCG CTTTAAAAGA	21120
GTTTGGTCAT GTTGACATTT GGGTAAATAA CGCCGGTGTG GAGGCTTCTT TTCCGACAA	21180
TGATATGCCA TTAAGAAGAT GGCAACGTGT CATTGATGTC AATTTAAACG GTGTCTTTTT	21240
GGGTTCTACG AC	21252

## (2) INFORMATION FOR SEQ ID NO: 95:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8726 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

GCATATCAAC AAGTCAAAGA ACGATATGAA ATTTTACAAA AAGAAGAAAA AGCCTTAAAG	60
CAACAACTGG CTTCTTATCA TGATtCTAT GAAACGTTTT GCTTGGCTGA AGAAGTCTTA	120
TTGGCGGTTA TCCAAAGAGA ACAAGCACGT CAACATTTGT TAAATGCTTC ATCATTTTTT	180
ATCTTGCAAA CGTGGATTCC AGTCGAAGAA AAAGCCGAGA TTTTGACAGC GATTGAAGAA	240
AAAGTTCCTA AGGACGAAAT AGCTCTTACT TTTGAAAATC CTACGAAGGC GGAAATTGAG	300
ACGGATATAC CAGTTAAACT AGCGAATAAT AAATTGGTGC AGCCGTTTGA AATGTTAACA	360
GAGATGTATA GCTTACCAAA ATATGAAGAA GTTGATCCGA CACCTGCGAT GATGCCGTTT	420
TATCTAGTCT TTTTGGTAT GATGGTAGCG GATATTGGCT ATGGCTTACT GATGCTACTC	480
TTGTCTATTA TTGCTTTGAA AGCATTGTT TTACCGAGAG GCATGAAACG TTTTGCCGAC	540

TTTTCTTGA	TTTTATCATT	TCCAACAATT	ATTTGGGGCT	TCATTTACGG	TTCGTTTTTC	600
GGTGCAGCAT	TACCGCCAAT	AATGTTTGGG	ATCAAATCAC	CATTTCCGAT	TTTATCAACG	660
ACCGAAGATG	TAAATACGAT	TCTTATTTTA	TCGGTTATTT	TTGGGTTTAT	CCAATTGGTT	720
GTCGGTTTAA	TGATTAATGG	GATTCAGTTA	TCAAAACAAA	AGCGCTATTT	AGATAGTATT	780
AACGAAAGTT	ACGCTTGGTT	AGGGATTTTG	TTTGGTTTAG	CATTGTTAGT	CGTTGGTAAA	840
TTAGTGGTTA	AAAATGAAGG	CTTATTTACA	GCAGGTGCGA	TTTtagCTAG	CCTTTCTGCC	900
ATTGCAATTA	TTGTGATTCC	AATGATTCAA	TCGAAGGCCA	AATTAAAAGG	CTTAGCCAAA	960
GGACTGTATG	GACTATACGG	TGTGACAGGT	TATGTCGGTG	ATTTAGTTAG	TTATACACGT	1020
TTGATGGCGT	TAGGAATTGC	TGGGGGAAGT	ATTGCTTCAG	CGTTTAATAT	GCTCGTAGAG	1080
TTCATGCCGC	CAGTTGCCCG	TTTTAGCGTG	GGCATCTTGC	TGTTAATTGT	TTTACATGCG	1140
TTAAACATAT	TCTTATCGTT	ATTAGGTGCT	TATGTTTCATG	GCGCACGTTT	GCAATATGTT	1200
GAGTTCCTTG	GAAAATTTTA	CACAGGCGGT	GGCCGGGCGT	TTAATCCGCT	AAAAACAAAA	1260
GAAAAGTATG	TCAATGTTGA	GAAAAAATAG	AAAATTATTT	ATTGGAGGAA	AATAGAGATG	1320
ATGGATTACT	TAATTAATAA	TAATGGTGGA	ATTGTTTTTG	CAGTATTAGG	AATGGCAATG	1380
GCGACAATTT	TTGCTGGGAT	TGGATCAGCC	AAAGGGGTG	GCTTTACTGG	GGAAGCAGCA	1440
GCAGCTTTAA	CGACGGAGCA	GCCAGAAAAA	TTTGGTCAAG	CGTTAATTTT	ACAGTTATTA	1500
CCAGGTACAC	AAGGACTATA	TGGTTTTGTT	ATTGCTTTCT	TAATTTATAT	TAAGTTAGGA	1560
AACGATATGA	GTATGGTTCA	AGGATTGAAC	TATTTTGTAG	CAGCTTTACC	AATTGCTTTT	1620
GCAGGATTGT	TCTCTGGGAT	TGCCCCAAGT	CGTGTAGCCG	CTGCGGGGAT	CCAAATTTTA	1680
GCGAAGAAAC	CAGAACATGC	AACAAAAGGA	ATTATCTATG	CTGCCATGGT	TGAAACTTAT	1740
GCCATCTTAG	GGTTCGTAAT	TTCCTTCTTA	CTTGTATTGA	ACGTCAAATA	ATTGAGGAGG	1800
CAGATGGATG	GATGCCATTG	AAAAATCAT	CAGTGAAATT	AAGCAGCAAG	GAAAGCAAGA	1860
AGTCGAAGCC	TATGTCACCA	GTGAACAAAC	GCGCATCGAT	CAAGAATTCC	AAGCAGCACA	1920
ACAAGAAATT	TTGCTGAAGC	AAGAACATGA	AATAGAAAAA	CGACAACAAC	AATTATTAAA	1980
AGAATTTAAA	CAACGCCAAC	AACGGCAAAC	ACTAGAAATT	CGTCAGGATG	CATTAAATAA	2040
AAAACAaGCA	TATCTTAACC	AaTTATTTGA	CGAaGTGGtC	TTAAAGATGA	GTGAATGGTC	2100
AGCGGAAGAA	TTTCAACAaT	TCATGAAAGC	ACAATTAGGT	TCACTTGAAT	TAACAGGAAA	2160
AGCAACCATT	CTTTTAGGCG	AGTATTCTCA	AACGAAAGTG	ACGCAAGAAT	GGCTGACCGC	2220
CCTTTCTGAT	GCTACTGTTC	AATGGGAGCT	TTCAGAGGAA	GTTGTTCCAA	AAGAAAGTGG	2280
CTTTATCGTT	GCAAAAGACG	GTTTGGATTA	TAATTTTCTC	TTTTCAGCAT	TAGTTGAAGA	2340
GATTCAAAAA	ACAGAAGGCT	TTAAAGTAGC	AGAAAACTT	TTCAGTTAAG	AAGGGAGGAC	2400
GTAGCATGAG	AAAACCAACG	TATCATCAAA	TTAATCCGCT	GATTGTTTTA	AAAGAAACGG	2460
AGCTGCTTTC	TGAGCAACAG	TTTCAACAAC	TTTTAGAAGC	GGAAACAGTT	GAAGACGTTA	2520

GAAATATGTT	AAAAAGTACG	GTTTATCAAC	CTTATTTAAC	CGAAACTTTT	GAAGAAAAAT	2580
TTGACTACCA	TTCTTCTGAA	GCGCTGGGAA	GTCTTTATCG	CTGGTTATAT	GAAATGGCTC	2640
CAGAACCAGC	TGTAGTAACG	TTGTATACGA	TGCGTTTTAC	CTTCCATAAC	TTGAAAATTT	2700
TGACGAAAGC	TGAGCGAACA	GGTAAAGATT	TTGATTATCT	TTATTTAGAA	GATGGTCGTT	2760
ATTCTCTAAA	CACAGTCAAA	AGTGCCATTC	ATACAAAAAA	CTCTTCCGAA	CTTGAGCCAG	2820
CATTATTAGA	AGTGATT TGT	GATGTTTTTG	CCTATCTGGA	AGAGGGCGGT	TTACCACAAG	2880
CAATTGATAT	CCTTTATGAT	CGAGCATTTT	TACAGCAACA	ACGAAAATTG	GCGGATGAAT	2940
TGGGCTATGA	AGAGTTAACG	AAGGAAGTCA	TCGCTTTTAT	TGATTTAACC	AATCTATCAA	3000
CGATGGCTAG	AGGGATTGTT	CAACATCAAA	ACAGCAACTT	TTTATCAACC	GTA CTTTCTA	3060
GCGCTGGCAG	TATCACAAAG	AAAGAACTGT	TGACCTATGC	GGAAAAGTCA	CTTGTAGAAT	3120
TTACAGCATT	TGTCAGAACA	ACGAATTACG	GACAACTATT	AGAAAAGATT	ATCAATAAAG	3180
AAACCA tGaa	TTAAATTTAT	TGGCTTTtGa	AAAAC TCAAA	GATGATTACC	TGACCTCAAt	3240
GTATGAaAAC	GGTCGAACGG	TCGCTTTtGG	ACCGCTCCCA	TTACTAGCCT	TTTTGAATGC	3300
GAAAGAGGTT	GAATGGAAAA	ATCTACGTTT	AATTTTGGTC	GGAAAACATA	GTGGGTTCCTC	3360
AAATGAAAAA	ATTCGAGAAA	GGATGCGAAA	AGTCAATGGG	GTATAAAATT	GGTGTCATTG	3420
GAGACAAAAA	TTCAATCTTA	CCATTTAAAC	TGTTTGGCTT	TGAGGTCCAT	CATGCCATTT	3480
CAGAAATGCA	AGTTAGAGAA	GCCATTGAAA	CTATGGCGAA	AAATAAGTTT	GGGGTCATTT	3540
TCATAACGGA	AGAAGCCTCT	ACTTTAGCAG	AAGAAACGAT	TGAACGATAC	AAAGAACAAG	3600
TGACACCAGC	GATTATTTTA	ATCCCAAGTC	ATAATGGTAC	AATCGGAATC	GGTTTAAGTG	3660
AGATTGAAAA	AAATGTTGAA	AAAGCAATTG	GACAAAATAT	TTTATAATGC	AAAGGAGAAT	3720
GAGCGTGCAA	ATTGGAAAAA	TTGTCAAAGT	TTCAGGTCCT	TTGATTTTAG	CTGAAAACAT	3780
GTCAGATGCT	AGTATCCAAG	ACATTTGTCA	TGTAGGAGAT	TTAGGCGTTA	TCGGAGAGAT	3840
TATTGAAATG	CGAGGCGACG	TCGCTTCGAT	TCAAGTATAT	GAAGAAACAA	CAGGCATTGG	3900
ACCAGGAGAA	CCAGTTATTT	CAACAGGAGA	ACCATTATCT	GTTGAATTAG	CCCCAGGTTT	3960
AATTGCCGAA	ATGTTTGATG	GATTCAACG	ACCATTGGAT	ACATTTCAAG	AAGTAACCCA	4020
CAGTAACTTT	TTAGGCCGTG	GCGTTAAAT	TGATGCGTTA	GATCGTGAGA	AAAAATGGAC	4080
GTTTGAACCA	ACTGTGGCAG	TTGGTGAAGA	AGTGTGCGCA	GGTGACATCG	TCGGTGTGGT	4140
TCAAGAAACA	CCGATTATTC	AACATAAAAT	TATGGTGCCT	TTCGGCGTTT	CAGGAACGAT	4200
TGCCGAAATT	AAAGCAGGTG	ACTTTGCCAT	TGATGAAACA	GTTTACTCAG	TGGAAACGGC	4260
TAAAGGAACG	GAAAGTTTTA	GCATGATGCA	AAAATGGCCC	GTTTCGGCGG	GACGTCCCAT	4320
TTTAGAAAAA	CTAAGTCCCA	AAGTACCGAT	GGTGACCGGA	CAACGCGTAA	TTGATACCTT	4380
TTTCCCAATT	ACGAAAGGCG	GAgcGGCAGC	AGTTcCAGGA	CCATTTGGCG	CTGGAAAAAC	4440
AGTCGTT CAG	CACCAAATTG	CTAAGTGGGC	CGATGTCGAC	TTAGTCGTTT	ACGTTGGTTG	4500

TGGGGAACGC	GGGAATGAAA	TGACAGATGT	TTTAAATGAA	TTTCCAGAAT	TAATTGACCC	4560
AACAACTGGT	GAGTCTTTGA	TGAATCGGAC	GATTTTAAAT	GCGAATACGT	CAAATATGCC	4620
GGTAGCGGCA	CGGGAAGCCT	CGATTTATAC	AGGGATTACC	ATTGCAGAAT	ATTTCCGTGA	4680
TATGGGTTAC	TCAGTCGCAA	TTATGGCGGA	TTCTACTTCT	CGTTGGGCAG	AAGCGTTACG	4740
AGAAATGAGT	GGTCGGTTAG	AAGAAATGCC	TGGTGATGAA	GGCTATCCAG	CCTATTTAGG	4800
TAGTCGCTTA	GCTGAATATT	ATGAACGAGC	AGGACAAGTC	ATCGCGTTAG	GAAAAGATCA	4860
TCGTGAAGGA	AGCATTACAG	CGATTAGTGC	GGTTTCGCCA	TCTGGTGGGG	ACATATCAGA	4920
ACCTGTCACT	CAAAATACGT	TACGCGTTGT	TAAAGTATTC	TGGGGCTTAG	ATTCCCAATT	4980
AGCACAAAAA	CGTCATTTTC	CTTCTATTAA	CTGGTTGCAA	AGTTATTCTC	TTTATTCCAC	5040
AGAAGTAGGG	CAATATTTAG	ACTTGGAATT	GCAAGGAAAC	TGGGCCgCTA	TGGTAGCTGA	5100
AGGGATGCGG	ATTTTACAAG	AGGAATCTCA	ATTGGAAGAA	ATTGTTCGCT	TGGTTGGGAT	5160
TGATTCCTTG	TCGGATAAAG	ACCGTTTAAC	gTTGGAAACA	GCCAAATCAT	TACGGGAAGA	5220
CTATTTGCAA	CAAAATGCTT	TTGATGACGT	GGATACGTTC	ACTTCTCGAA	CAAAACAAGC	5280
GAAAAATGTTG	CAATTGATTC	TAACTTTGG	TGAAGAAGGT	CAAAAAGCCT	TAAGTTTAGG	5340
CACTTATTTT	TCTGAGTTAA	TGGCGGGAAC	AGTTGAAATC	CGCGATCGCA	TTGCTCGTAG	5400
CAAGTATTTA	CCAGAAGAAG	AATTAGAAAA	ATTGGATCGT	TTACAAGCAG	AAATTAAAAC	5460
AACGATAAAA	GAAATCATTG	CTGAAGGAGG	AATGACGAAT	GATTAAAGAA	TATCGTACAA	5520
TCAATGAAGT	CGTTGGTCCT	CTGATGATTG	TTGAAAAAGT	GGCAGGCGTA	AAGTACGAAG	5580
AATTAATTGA	AGTACGCATG	CAAAATGGCG	AAATTCGCCA	AGGGCAAGTT	TTAGAAATCA	5640
ATGGAGATAA	AGCGATGGTC	CAAATTTTGT	AAGGAACGAG	TAACATCAAT	ATTCGTGATT	5700
CAAAAGTTCTG	CTTCTGGGA	CATCCTTTAG	AATTAGGGGT	TTCGCCAGAT	ATGATGGGGC	5760
GCGTTTTTGA	CGGCTTAGGT	CGCTTAAAAG	ATAATGGACC	AGAATTATTA	CCTGAGAAAA	5820
AATTAGATAT	TAACGGCGAA	GTTATCAATC	CAGTTGCTCG	TGATTATCCC	GATGAGTTCA	5880
TCCAAACAGG	GATTTACGCG	ATTGACCATT	TAAATACCTT	AGTTCGTGGT	CAAAAATTAC	5940
CTGTTT̄TTTC	AGCATCTGGC	TTACCTCATA	AAGAATTAGC	GGCACAAATT	GCCAGACAAG	6000
CCAACGTGTT	AAATAGTGaA	GAAGAATTTG	CCGTAGTTTT	TGCGGCCATT	GGGATTACCT	6060
TTGAAGAAGC	GGAATATTTT	ATGGAAGATT	TCCGTCAAAC	AGGCGCAATT	GATCGTTCAG	6120
TCTTGTTTAT	GAACCTAGCG	AATGATCCaG	CCATTGAACG	GATTGCAACA	CCTAGAATGG	6180
CTTTGACAGC	CGCTGAATAT	TTAGCTTATG	AAAAGGGCAT	GCATGTCTTA	GTTATCATGA	6240
CGGATATGAC	AAATTATTGC	GAAgCGTTGC	GAGAAATTTT	AGCAGCACGC	CGTGAAGTTC	6300
CAGGACGTCG	TGGTTACCCA	GGTTATCTTT	ACACGAACTT	AGCAACGTTG	TATGAACGGG	6360
CAGGCCGAAT	TCGTGGCTCA	AAAGGTTCCG	TAACACAGAT	TCCTATTTTA	ACAATGCCAG	6420
AGGAGGATAA	AACACATCCA	ATTCCCGATT	TAACGGCTA	TATTACAGAA	GGGCAAATTA	6480

TCTTGTCCCG	GGA	ACTATAT	AAGAGCGGTA	TCCA	ACCACC	AATTG	ATGTA	TTGCC	ATCAC	6540		
TTTCCCGTCT	TAAAG	ACAAA	GGA	ACTGGCG	AAGG	CAAAAC	GCGCGGGGAT	CATG	CaGCGA	6600		
CGATGAATCA	ATTG	TCTCA	GCCT	ATGCAC	AAGG	GAAACA	AGCCAAAGAA	TTAG	CTGTCA	6660		
TTTTAGGAGA	ATCAG	CTCTT	TCCG	ATGTCG	ATAAA	ATTTA	CGCAGCTTTT	GCCCA	ACGTT	6720		
TTGAAGAAGA	ATATG	TCAAT	CAAGG	ATTCG	ATACA	AACCG	TTCGATTGAA	GAAAC	GTTAG	6780		
ATCTTGGGTG	GGA	ACTATTA	AGTAT	GCTAC	CGAGA	ACAGA	ACTAAAACGA	ATTAA	AGAGG	6840		
ATATGCTCGA	TCAAT	ATTTA	ACTGA	AGGGA	AGTAG	AAAAA	TGGCTCGATT	AAATG	TTAAT	6900		
CCAACCCGTA	TGGA	ACTTTC	CCG	ATTGAAA	AAAC	AGTTAA	CTACGGCCAC	GAGAG	GTCAT	6960		
AAGTTGCTGA	AAG	ACAAACA	AGAT	GAATTG	ATGCG	ACGGT	TTATTGCGCT	AGTCA	AAGAA	7020		
AACAATGAGT	TACG	TATTCA	GGT	TGAACAA	GAAGT	GACGG	ATGCGTTATC	AAATTT	TGTA	7080		
TTGGCAAAACG	CAACG	TTGAA	CGA	AGCCTTT	ATTGA	AGAGT	TAGTGGCAAT	TCCAG	CTGAA	7140		
AAAGTTGAAT	TAGAA	ATTAT	TGA	ACAGAAC	ATCTT	AAGTG	TGCCAGTTCC	TAAAT	GATT	7200		
TTTGATTACG	ATGA	ATCGGT	TCA	AGAAGCA	CCCTT	AGATT	ATGGCTACGT	GAATT	CAAAC	7260		
AGCGAGCTTG	ATCA	AGCTTT	TGCT	AAAAATT	TCAAG	CATTT	TACCGAAATT	ATTAG	CACTT	7320		
GCAAATGTTG	AAAAG	ACGTG	CCA	ATTATTA	TCAAA	AGAAA	TTGAAAAAAC	GCGT	CGTCGT	7380		
GTCAACGCCT	TAGAG	TATAT	GACG	ATTCCA	CAATT	rGARg	AAaCa	aTTTA	CTmCa	TTmCa	7440	
ATGAAATTwG	rAGr	AAATGA	ACGT	TGGcGAA	ATTAC	ACGTC	TAATTA	AAAT	TAAAAGCATG	7500		
AACA	AgAAAA	TTAGT	GGTTA	GGA	AGAACGA	GTGCG	AAcCa	GTTGGT	AGCT	TTCGT	GCTCG	7560
TTTTTAATAA	AAAAG	GGAGT	GGAAAA	AATG	AAAA	ATGAAG	AACGCCGCAA	AGCG	ATTGCA	7620		
CTGAATTGTC	AAAA	ATACGA	AAGT	GACTAT	GCTCG	CTTG	TGGAACCAAT	CAATGa	GCTA	7680		
CTGT	TAAACT	TGGG	AGCGGC	TATTT	CAGAA	GAAGCCgCGA	AcAAATCATT	CTGA	ATGTAA	7740		
AACGATATCA	TCACG	CGGTG	AAAT	ATCTTC	CTGAGT	GTCA	TTTAGATGAA	AGTA	ACCAAT	7800		
TTATCGAAGA	TGGG	CTTGAA	GCTCT	TAAAA	AGGG	AGATTT	AGGCAACGGT	GCGT	TACAAC	7860		
TATTTGGGGC	CGGCT	TAAAC	TTTG	CTAGTT	TTGC	AGCAAA	AGCATAAGGA	ACCA	AAAAAAA	7920		
TTGATGCCCA	CCAA	ATGTTG	GCAG	AACGTT	TtAC	GAAGTT	GTTGTCTGTA	CaAAC	GCGATA	7980		
ATGATAaCAA	ACA	ATAAGrA	AAAT	ACCCTk	GGTTT	TATTAA	GTCGACTTAG	TAAAC	CAAGG	8040		
syATTTTTTTT	AACTG	TATT	GATTAG	GGGTA	TTCAG	ACACC	TGTAATTGCC	ATAAC	TATC	8100		
TGTC	ACTGAA	GCA	ATAAACG	CAGCA	CTCTT	AGCATTATCA	GGTCCTTTAG	TAAAA	ATTGC	8160		
TAAAGCAAAG	GGGT	GAGGTG	TTTCT	AAAAAT	CCCC	ATATCA	TGAATAAATT	CATCA	TAGGA	8220		
ACCAATTTTA	TGAG	CAACTT	TTCCT	TGAGT	TGTGGG	AGTT	TCCATTCTTT	CATG	GAAAAC	8280		
ACTTTGTTTC	ATAGA	ATCAT	AAAT	CGCTTG	GTAAT	CTGGT	TTTGTTGCTT	TTTC	AGTATA	8340		
TAATTTTTTGT	AGA	ATAACAA	GCG	CATCTTC	AGA	AGAAAAAT	TGTGGCTCTT	CaATC	GAGG	8400		
TGTCTTGTGC	AAATA	ACGCT	GATAC	ATTTTC	TCGTTTTGCT	TTTG	CATCAC	CACCA	ATGT	8460		

631

GTCATAGAGC ATATTTTATAG CGATATTATC AGAATAGGTG ATGGCATATT CTTGTAACGT	8520
TTTTAATGGA TATTCTGGTT GGATATTATA GGCAATGATC CCAGTTCCT TCTTCATAAT	8580
CTTCTTCGGC GTTATAGGGA ATCAGGTCTG TCCATTTTTT TTGACCAGAA GCAACCGTGT	8640
CGGCAACTAG CATCGTCAGG GGCACTTTAA TAGTACTAGC TGTATAAAAT TCATGTTGGC	8700
CATTACGGA ACGCTGTTTCG CCAGTT	8726

(2) INFORMATION FOR SEQ ID NO: 96:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2531 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

GTAAGACTT TTCACAAGTA TCAAAGAAAT CCGTCGTAA TCTTTACTAT TCGACAAACG	60
AACACCCTCT TGCTCAGCAA GAGGGTGTTT TTTGTGTAA TTAAGTGGAA AATCAACATT	120
GACTTTTTTT GAAGTCACAT AAGTGTGAAA ATTAGGTTTT TGGACAATGA TTCAGAAATC	180
TTTTGAAAAT TGAAGGTGTG AGGAAATTTA GAAAGATAAA AAGAGGGAGT ACTGAAAACC	240
GAAATTCAG TTAAGTGCTT GGATCAAGAA AAAAAACTA CGTTAKTTTG TCAATATCAG	300
TTGAAATGG AGAAGAAGCA GAATGGGGTG ATAGTATTGG CAAATTAATT ATTTAAGGTC	360
ATCGTTACAG GATTTTTGTT TTTTATATyC aATCmAAAAG AAATAAATAC AAAGaACACA	420
GATAAGTATT TTAGCAAGTA ACCAATAAAT AGCAGGACTG TTGGTGCTAT AAGAATACAA	480
GAAAATAGTA GTAAACTTAA TTGAGAATCG TTTGCTATTT AAATAGTGTT CTTTTCCTTT	540
TAATATATAT ACGATAAAAT AAGTATATTA ATTGTAAaAG GACATGATAA TTATGGAAGA	600
ATTAGAATGG AATwTTTTT aTGTTTGCTA TAGTTTTGAG GGTAGTGATT CTTTTAATAA	660
CTGTCAAGAT AGTAGATTAC TTTGnGAATA CTCCTGAAGT ACAAATACAG AGAAACTTC	720
CTGAGTTTGA AAAAGTGCCA CCTTATGATC CAAGAGATAA TTCATTTCTG ATACAAATTA	780
TTCTTTAAGT AATACAACGA CTAAGTTTAT TTACCAAACC ATTACAATAA GAAAAGGTGC	840
AAGTATACCT AATTACTATT GGTATGATGA TGGTCAATTT AGAGGAAATC TAAAACTTCA	900
ACGTTGGCAA GACTACGGAG AAGATCCAAA CATAAGAGGA TATATCTGTA CTTATGGAGG	960
CTACGTTACC AATGGTTCAA GCCCGGCGCA AAGAATTGTC GATATTAATG AACAATAGCT	1020
ATTCTAACTT TATGTAATAG GAGGGAGTAA TCATGAAAAA AATCGGGTAT TTTAGTTGTA	1080
TTATTTTTTT CATGTTTTTG GTAGGTTGTA GTAATAACAA AAAAGAAAAC GGCAATCTTT	1140
TGAATGCCAG TTCGTTTCCT TTAATACTCA CCACGATTAT TGAAAAAGAA GAAGACCTAA	1200
CGAAAGGTTT AATTTTTTTC AACAAGGATA AAACCATGAC GCTTGAAAAA GAATATTTAG	1260
TTAATCCCAA TAATGAAGAC AAAAAAAAAA CAAGTAGAAC AGAAAAAAG GTATATAAAA	1320
ATATTAAAAA ACAAGAAAAT AAAGAGAGCT ATGAAATTAT AGGTCAATTG GACAAAAAAA	1380

CGAAAAAAT AGAGTTTAAA AAAGTTGATG AAGGTAAACG TATATCTGAT GCAGAAGGTA	1440
ATGTGTATGG TGATTTTGGT GGTAAATAGG ATAGTTAAGT CATAAAAAAG AAGGCTAAAA	1500
ATTTTATTT TTTTCGTCTC TATACTTTCT GATGTTAGTT TTTTACCAA CATCAGAAAG	1560
TATAGAGACG TTTTGTGTA ATTAAACAAA AAATCAACAT TGACTAAGAC CAGAAATACT	1620
TCCTTTAGTG CTCCGAGTTG CTCTAATATA AGACACCAA TTGATAACCA TTTGCTTGAT	1680
AAACTCTAT TAATTGTGGC AATGCTTGT TTGTTAATTC CTTATCTACT GAATCATGCA	1740
TTAAGACTAC TCTAATATTA TAATCCGGAT AAACCTCAAG TGAATGTTGA TGGAAAGCGA	1800
GCATTTCTGC GACTGTAGTA GGCTGTCTGT CTAACGGCTC CGCATCTCCA ACCATCGCAT	1860
TCCAGTCAAT CCAATGGATG CCAAGTTGTT TGAATAATTC ATCAGATTTT TCTGTACCTC	1920
CCCATGACAT ATGACCTCCT GGGTATCGCC AGAGAGTTGT TTGAAATGTT TCGCCAATA	1980
TTTTTTTAAAG AGAGTTCTCC ATGGTTGCTA TTTCCGCTTG TATTTCTTTA GTGTTGACAA	2040
TCCCGTTTGG ATATAACGTT GCATAATCAT GAGTTGAAGT ATGAAAGCCT ATTGAATGTC	2100
CTTCTGCAAC TTCGCGTTTC ACTATCTTTT GATTTTCGGA AGTTAGGGTA TTCCTACAA	2160
GAAAAAAGT TCGTGGACG TTATACTTTT TTAATGTATC TAGGACTTGT GGGGTtATTT	2220
GATTATTAAC GCCATCATCA AAGGTTAAGA ATACTAATTT TTCACCTTTG TAAGGTTCTT	2280
CTCCAGTTAT ATATTTTCTG ATTTTGTCTG CTGAGTAAGC ATATTTTTTTT GATGAAATAG	2340
TCATGTGTGT TGATTCCTTT TCTAACTTGG TCGGGGAATC TTTTTTTGCA GTTGAATTTT	2400
TAGTCGATGT CTGCGTAGTA TCATCATTTT TACTTGTTAT CTTGGGTTGA GTTAATAGGA	2460
ACGTTTGAAT GAAAACAAGA ACAATGAGGA TTAACGCACC TAAATnATAC ATACCCGTnT	2520
ATTTTGAATT T	2531

## (2) INFORMATION FOR SEQ ID NO: 97:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2717 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

ACCGCCGcG CCAArGAtGa AcGAAAATAT CTcGAATCCA CAAAACCAC TACTGCTAAA	60
TCCGCCGCCA CCGAAACCGC CAAAGCCaCC aGCGCCGcCA CCGwAATTAG GATCAGTTCC	120
TGCATGGCCA TATTGGTCAT AGGCCGCTTT TTTCTGTGGA TCACTTAATA CTTCATACGC	180
TTCAGAAACT TCTTTAAAT TTTCTTCCGC ATCGGCTTCT TTGTTAATAT CTGGATGGTA	240
CTTTTTGGAA AGTTTGCGGT ATGCTTTTTT AATTTATCA TCTGAAGCAC CTTTGCTAA	300
TCCTAGCACT TCATAATAAT CTCGTTTCGT CGCCATTGAC TTCCCTCAA CTTTCTTCAA	360
ACTTTACAAT GTGTTTAGAA CCAGCCAAAT TCATCGATTG CTTTCGGTCT CTCTGGTTGG	420



ATTATCTGTC ACACCTTGAA TCAGTGTTAC TAGCTGTGAA TAGGTATTTT CAGCTTGTTT	480
AACATCGTTG TAATCATACC ACAGTTCTTT CAAAACCTAA ATAGTTTCTT GCTAACTTGA	540
TTGTGGTTGA TTTTCTCACA AAAGACAAGG GACTGGGACA TTAGTCATTG TACTAATGAT	600
CCAGTCCCTA AAATCaATTG TTATTGATTA TTTGTCAATCA CCATTTACTT CTTCAAAATC	660
AGCATCAACA ACATCATCAG CGCCACCTtG gAGCAGCTTC AGGgATTTTC TTGTGCTTGT	720
TGTTGCGCat TGgTTCGTAA AGTTTTACTG TTAAGTTTTG GACGATTTCA TTTAATGAAT	780
CACGTTTTGC TTTCATTTGT TCGATATCGT TTGCTTCGAT GGCTGCTTTT AATTCATCAC	840
GCGCATCTTC TGCTTTTTTC ACTTCTTCAG CATCTACTTT GCCTTCTAAT TCTTTCAATG	900
TTTTATCAAC AGTGAATAAT AAAGCATCTG CATCGTTACG TAAGTCAACT TCTTCTTTAC	960
GTTGTTTATC TGCTTCAGCG TTTGCTTCAG CATCTTTTCA CATACTTCG ATTTTCATCAT	1020
CAGATAAACC TGAAGAAGAT TTAATCGTGA TTGTTTGTTT TTTTGTAGTG CCTAAGTCTT	1080
TCGCACGAAC ATTTACAATC CCGTTTTTAT CAATGTCAA GCTTACTTCA ATTTGTGGCA	1140
CACCACGAGG AGCTGCAGGA ATATCTGTTA ATTGGAAACG ACCTAATGTT TTATTATCAG	1200
CGGCCATtGG ACGTTCACCT TGTA AACCGT GAATATCAAC AGCTGGTTGA TtGTCAGCAG	1260
CAGTTGAGAA TACTTGTGAT TTACwTgtTG GAATCGTAGT GTTACGATCa ATTAATTTTG	1320
TAAATAcGCC ACCCaTTGTT TCAATCCCTA ATGACAACGG TGTTACGTCT AATAAgACAA	1380
CGTCTTTAAC ATCACCAGTG ATTACACCAC CTTGGATTGC AGCACCCATT GCTACTcTTC	1440
GTCAGGGTTA ACTGATTTGT TTGGTTCTTT GTTTGTCTTCT TTGCGAACAG CTTCAACTAC	1500
AGCTGGAATA CGTGTTGAAC CACCAACTAG GATAACTTCG TCGATTTTCA AAGGATTTAA	1560
TCCAGCATCT TTTAAGGCTT GACGTACTGG TACTTTTGTA CGTTCAACTA AATCACTTGT	1620
TAATTCATCA AATTTTGCAC GAGTTAAGTT CATTCTAAG TGCAATGGTC CTGcTTGCGC	1680
AGCAGTAATA AATGGCAAGC TGATTTGTGT GCTTGTTACA CCTGATAAGT CTTTTTTCGC	1740
TTTTTCAGCA GCATCTTTCA AACGTTGTAA TGCCATTTTA TCGTTAGCTA AGTCAATGCC	1800
ATTTTCTTTT FTAAATTCTG CAACCATGTA GTCAATGATC TTATTATCAA AGTCATCCCC	1860
ACCTAGGTTG TTATCACCGG CTGTTGATAA TACATCGAAA ACGCCATCGC CTAATTCAAG	1920
GATTGACACG TCGAATGTAC CACCACCAAG GTCAAAGACA AGAATTTTTT CATCTTTGTC	1980
TGTTTTATCT AAGCCATAAG CTAAAGCTGC TGCTGTTGGT TCGTTAACGA TCCGTTCTAC	2040
TTCTAAACCA GCAATTTTAC CAGCATCTTT TGTTGCTTGA CGTTGCGCAT CGTTGAAATA	2100
TGCAGGAAC GTAAATACTG CTTTTTCAAC TTTTTCACCT AAATAGTCTT CAGCAAAACC	2160
TTTTAAATAT TGTAAGATCA TTGCAGAAAC TTCTTGTTGGT GTATATGATT TTCCTTCAAC	2220
ATCAACTTTG TAGCCAGCTT CACCCATGTG ACGTTTAATA GATGAAATCG TGTTAGGATT	2280
TGTAACAGCT TGACGTTTTG CTACTTCGCC CACTTgAATT TCACCGTTTT TGAATGAAAC	2340
GACCGAAGgA GTTGTGCGGT TACCTTCTGG GTTTGCAATA ATTTTCgetT CGCCGCCTTC	2400

634

TAGTACTGCG ACAGCTGAGT TTGTTGTTCC tAAGTCAATA CQAATAATTT TACTCATAGT	2460
GAATATCTCC AATCTGTATT AGTTTCTAT AATAATATAT TTTTAATTAA CGTTTGTTAT	2520
TGTGCAACAA TGACCATTGA TGGTCGTAAC ACACsATCTT GTAATTTGTA CCCTTTTTGT	2580
AGAACTTCAA CAATGGTATC TGCTGGGGTT TCTTCAGAAG CAGGGACTGT TTGAACGGCT	2640
TGATGTAAAT TTGGATCAAA GGTTCACCCA TTGCTGGGAT TTCTTCAATT CCTTCTTCTT	2700
TTAAAGCCAC AGTAATA	2717

## (2) INFORMATION FOR SEQ ID NO: 98:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16870 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

GAATTTAGAA ATGAGGACTA CTGATGATTT TATTACAAGC AAATCAAGTT GCCCGGCATT	60
TTGGCTCGGA AACATTGTTT GAAAACATAC ATTTAGAAAT TGCAACAAAA AGTCGGATTG	120
CCTTAGTTGG TCGTAATGGT GCTGGAAAAT CGACTTTTTT AAAAAATCATT GCAGGCATTG	180
ATGCTCCCGA TAGCGGAACC ATTGCCAAAA ATaAAACTGC TACGTTAGGT TATTTAGCTC	240
AAAATACCGG CTTAGAATCA GATAAAACCG TTTGGGAAGA AATGACAAAA GCCTTTGCTG	300
ACATCCTAGA AATGGAACAq CGTATGCGAG aATTAGAAAC TAAAATTAGT GAAATGGAGC	360
CAACCACTTC CGTTTATGAA GGAATTTTAA AAGAGTACGA TCAATTGCAA CATACTTTT	420
CTGAAAAAAA TGGCTACGGC TATGAAAATG AAATTCGCTC AGTCCTTCAC GGCTTTGGCT	480
TTGATGAATC CTTTACACG AAAGATATTC AAACCTTATC TGGTGGTCAA AAAACCCGGC	540
TTGCATTAGC GAGAATGCTT TTACAAAAAC CAGACATTTT AATTCTGGAC GAGCCTACAA	600
ACCACTTAGA TATCGAGACG CTTtCTTGGC TGGAACTTA TTTGCCAAGT TATGCCGGCG	660
CCCTATTAAT TGTTTCCAC GATCGtTATT TTTTAGATAA GGTAGTTAAT GAAGTTTATG	720
AACTGAGTCG CAAAAAATG ACTCACTACA AAGGaAACTA TTCCAATAC TTAGAGTTAA	780
AAGCAGAACA ATTAGCCAGT GAATGGAAG CGTATGAAAA GCAACAAGAA GAAATCAATA	840
AGTTAGAAGA TTTCGTTGCC AAAAACTGG TTCGTGCATC TACAACGAAA CGTGCACAAA	900
GTCGCCGAAA AGTATTAGAA AAAATGGACC GTTTAGACCG ACCTCAAGGA GATGAAAAAT	960
CGGCGCATT TCTTTTCGAT AGTGAAAAAG TCTCGGAAA TGTGTTTTTA CAAGTCGAAG	1020
ATGCCGCCAT TGGTTACGAC CAAGAACATA TTTTATCCGA ACCTATTCAC TTGGATATTC	1080
GTCGCAAAGA AGCCATTGCC TTAGTCGGAC CGAACGGAAT TGGTAAATCC ACTCTCTTGA	1140
AATCAATTAT TGACCGCATT CCTTTCATTA AAGGAAGTAA AACTTTTGGC ACCAATGTTT	1200
CTGTAGGTTA CTATGACCAA GAGCAAGCCA ATTTACATGG CAATAAAACG GTCTTAGCGG	1260
AATTATGGGA TGAACACCCA ACCACACCTG AAAAAGAGAT TCGAAGTATT TTAGGCGGCT	1320

TTCTCTTCAG TGGAGACGAT GTTGAAAAAA CGATTCCTTT ATTAAGCGGT GGCGAAAAAG	1380
CCCGTGTGGC ATTAGCAAAA CTAGCGATGG ATCGTGACAA TTTCTTGATT CTCGATGAGC	1440
CAACCAATCA CTTGGATATC GATAATAAAG AAGTTTTAGA AAATGCGCTG ATTGATTATG	1500
AAGGAACCAT CCTCTTCGTT TCCCATGACC GTTACTTTAT CAATCGAATT GCAACAAAAG	1560
TTGTTGAGCT TTCTGAAAAA GGCAGCAAAC TTTATTTAGG CGACTATGAT TATTATTTAG	1620
AAAAGAAACA aGAGGAAGAA GAAATCGCTG CCCTCTTAGC TAATGAAGAA GCGGCGAAAA	1680
AACCCGAACC AGTTACAGCC AAAAATACCT TTTATCAAAA CAAGGAGCAA CAAAAATTAC	1740
TCCGTACTTT GCAAAGAAAA ATAACACAAG TCGAAGAAAA TCTTGCTCAG TTAGATACGA	1800
CAATTGCACA ATTAGAAGCA CAAATGAGTC AACCAGACAT TCTAGAAAAT CATGTTGAAT	1860
TGCTGGCTTT AAACCAACAA TTAGATGAAG TTCGTCAGCA ACAAGATGAA CTAATTGAGC	1920
AATGGGAAAA TTTCAGTTTA GAATTAGAAG AAATGGAAAA TAACAATTAA CGGAGGGATT	1980
CACTATGGCT TCCAAAAAAA TCAGTTTAAAC GAaGCCCCAA TATCAACAAA TCGCCGTGTA	2040
TGTTGCGGAA AAAATTGCAG AAGGCAAATT ACATGTTGGA GATAAAATCC ATGCCCGTTC	2100
CACCCTAGCT AACCAATACC AAGTGTCGCC AGAAACAGCT CGCAAAGCCA TTATCGTTTT	2160
AGTTGATTTA GAAATCGTTA AAGCCAAACA CGGCAGCGGC TTCTATGTGG CATCAAAAGA	2220
AAAAGCGCAG GACTTTGTGA CACAATATCA AGACGTTCAA ACGATTGCTG AAATTAAAGA	2280
AGAATTGCTA GATAGCGTAG cGAAACAAA AGAAGAACTT ACTCATTTTT CTAGCATTCT	2340
AGATACTCTT GTTGAACAAA CAAAGCGTTT TGATTCTTTC AATCCAATGA ATCCCTATTC	2400
ATTAGTTTTA ACAGAAGAAG CTGCTTATCT TGAGGCGACG ATTAGCGAAA TGAATTTTTG	2460
GCAAAACACT TCGGCAACGA TTATCGCCAT CAAACACAAA GAAGAGCTTC TGGTCTCACC	2520
TGGTCCATAT GCAAAGATTT CGTTGAACGA CACACTTTAT TTTGTTGGAC ATGACGAATC	2580
AACCTTACAA CGTGTCCAAA ATTTCTTTTA TCCTTAAACA AAAAAGTGAC TACTTATCTT	2640
GCGGTAAATA GTCACTTTTT TTATTTATTA TTTCTATTTT TCATAAAAC CCACTATAAA	2700
ACCTAGTTAA AATACTGTTT TTCTTCATTT TCAGAAAGAT TTTCATCTAA TTTGACAAAG	2760
TGACAACTTT ATAGTAGTCT GATGTAGTCA CTTTTTTTAA TGAGGAGGAA CAATGATTTG	2820
CCAAAAGTAA AAgtAAATCA CCTTACCAAG ATCTTTGGtA AGAAAAACCA ACCGGCACTT	2880
GAAATGATTC GTGCCaACAA AAGCAAAACA GAAATTTTAG AAAAAACGGG TGCTACAGTC	2940
GGTGTTTATG ATGTAAATTT CGAAGTTGAA GAAGGCGAAA TTTTCGTTAT TATGGGGTTA	3000
TCAGGAAGTG GGAAATCAAC ATTAATTCGT TTATTAAATC GTTTGATTGA ACCAACTTCA	3060
GGAAATATTT ATATTGATGG ACAAGATATT TCTTCTTTAG ATAAAGAAGG ATTACGTGAA	3120
GTACGTCGAA ATAAATGAG CATGGTTTTT CAAAACTTTG GTCTTTTCCC CCATCGAACT	3180
ATCTTAGAAA ATACGGAATA TGGTTTAGAA ATTCGTGGTG TTCCTAAAGA AGAACGCCAA	3240
GCAAAAGCTG AAAAAGCCTT AGAAAATTCT AGCTTAATTG CTTTTAAAGA CCAATTACCA	3300

AGTCAATTAT CTGGTGGGAT GCAACAACGT GTCGGCCTTG CTCGTGCCTT GGCTAACGAT	3360
CCAGAAATTT TATTAATGGA CGAAGCGTTC TCTGCTCTTG ATCCGTTAAT TCGTCGTGAA	3420
ATGCAAGACG AACTATTAGA TTTACAAGAA AATGTCAAAA AAACCATTAT CTTTATCACA	3480
CATGATTTGA ATGAAGCTTT ACGAATCGGC GACCGGATTG CTTTGATGAA AGATGGTCAA	3540
ATTATGCAAA TTGGGACTGG AGAAGAAATT TTAACCAACC CAGCCAACGA ATATGTGCGA	3600
ACCTTCGTGG AAGATGTGGA TCGTTCAAAA GTCTTAACTG CCCAAAATAT TATGGTACCT	3660
GCTTTAACTA CTAACATTGA AATTGATGGC CCAACTGTTG CTTTAAAACG GATGCGTCAA	3720
GAAGAAGTGA GTATGTTACT TGCCGTGGAT AAAAAACGCC AACTAAAAGG CGTTGTACGC	3780
GCAGAAAAAG CTTTAGAAGC GCGTAAAAAT GGGACTTCTC TTGTAGAATG TGTGGATCCA	3840
GAAATTCAAA CAATCGACAA AGATATGTTA GTGAATGATA TTTTCCCACT GATTTATGAT	3900
GCACAAACGC CTCTTGCCGT GACTGATAAC GGCAAACTAT TAGGCGTGGT CATCCGTGGT	3960
AGCGTACTTG AAGCACTTGC AGAAACAGAG GTGAACGAAC ATGAATAAGT ATCAATTACC	4020
TGTCGCGTCT TGGGTTGAAA GTTTTACTGA TTGGTTAACT TCAACATTTG CTGGATTATT	4080
CAGCTTTTTA CAAACAATTG GTCAAAGTGT AATGGACAAC ATTACTGCCT TACTAACAGC	4140
CGTACCACCA CTTGTGCTGA TCGTGCTATT AACGATTGCT GCTTCTTTA TTTCTAATAA	4200
AAAAATTGGC TTAAGTTTAT TCACGTTTAT TGGCTTAATG TTTATCTATA ATCAAACTT	4260
ATGGAACGAT TTAATGAGTA CCGTTACGTT AGTTTTACTT TCTAGTGTTA TTCCATTAT	4320
CATTGGTGTG CCACTAGGGA TTTTAATGGC GAAAAGCGAA AAAGCCAAGA GTATCATTAC	4380
ACCGATTTTG GACTTCATGC AAACAATGCC TGGTTTCGTT TACTTGATTG CAGCCGTTGC	4440
ATTCTTCGGA ATTGGTATGG TTCCTGGGGT CTTCGCCTCT GTTATCTTCG CTTTACCACC	4500
AACCGTACGC TTTACGAACT TAGGAATTCG ACAAGTGCCA AATGAACTTG TTGAAGCAGC	4560
TGATTCTTAC GGTAGTACTG GCTGGCAAAA ATTATTCAAA TTAGAATTAC CTTTAGCCAA	4620
AAACACGATT ATGGCTGGTG TTAACCAAAC AACCATGCTA GCCCTATCAA TGGTAGTTAT	4680
TGCTTCAATG ATTGGTGC GCCTGGATTAGG ACGTGGCGTC CTTTCCGCTT TACAACGTGC	4740
ACAAGTCGGG AACGGTTTCG TAAATGGTGT GGCTTTAGTT ATTTTAGCGA TTATCATTGA	4800
CCGTTTTACT CAACATTTAA ATAATAAAAA AGCAGCTCCT AAAGCAGCTG GTGCCACCTC	4860
TAAAAAGAAA AAATACGGGA TTATTGCAGC GGTGTGTGTA ATTGTCGCTG GCTTAGTTGG	4920
TGCTTCAATT TTTACAACAA CAAATGATAA GCAAATTTCC CTTTCTTATG TGGAATGGGA	4980
TACCGAAGTA GCCTCCACAC ATGTGGTTGC CGAAGTCTTA AAAGATATGG GGTATGATGT	5040
TAAAACGACC CCTTTAGATA ACGCAATCAT GTGGGAATCT GTAGCTAAAG GTGAAACAGA	5100
TGCCATGGTT GGTGCGTGGC TGCCAGGAAC TCACGCAGAG CAATATAAAC AATATAAAGA	5160
TAAATTAGAC GACCTTGGTG AAAACCTTAA AGGCGCGAAG TTAGGAATTG TTGTCCCTTC	5220
TTATATGGAT GTTGATTCTA TCGAAGATTT ATCTGATCAA GCTGGGAAAA AGATTACTGG	5280

AATTGAACCT GGGGCTGGTG TTGTTGCAGC AGCAGAGAAA ACTAAAGAAG CTTATCCTAA	5340
TTTGAAAGAT TGGTCTGTTG AAACCTCTTC ATCAGGTGCC ATGACTGTCG CCTTAGGACA	5400
AGCAATTAAA AACAAACGAAG ACATCGTTAT TACTGGCTGG TCTCCTCACT GGATGTTTGC	5460
AAAATATGAT TTAAAATATT TAGCTGATCC AAAAGGCACT ATGGGCGGCG AAGAAGCGAT	5520
TCACACAATG GCTCGCCAAG GCCTAAAAGA AGATCAACCA GAAGCTTACA AAGTGTTAGA	5580
TAATTTCCAT TGGACCACTA AGGATATGGA ATCTGTCATG TTAGAAATTA ACGAAGGCAA	5640
AGACCCACAA GAAGCTGCAC GTGATTGGGT TGATTACAT AAAGACCAAG TAGCAGAATG	5700
GnAAAAATAA TTCTTAACAA ATGACTTCCC CATCTATTAA AAGrGGGCGT TGAACAAATT	5760
GTTCAACGCC CTCTTTTTTC GCAAAAAAAA CCACTCAGTT GGGTAGACTG AGTGGAAGG	5820
AGTTATTTTT ACTTGAGGAG TAAAAATGAA AAGTGTTTAT GTTGGGTGTT TTGTTGGTAT	5880
GCTTATATAT TAACGAGAAC CTGTGAATAA TTTGTGACGA AAACCTTGA AATTGCTAC	5940
GAATTACTTA AAAAAGTTTA AATGTGTCTC ACCCTAGTTA ATCGATACAG TGATTATTTT	6000
TTTTTAGCAT TTTGGATGGC TTTAAATAAT GTTGAAAATT GTCCTTTGAA AGACATCACA	6060
TCATCAAAGA TATCAACACC ATCGTTTAAC ATTTTGACGG TAACCCAGCC AAGCGCTTCC	6120
GTAATTGCTA CCGCAACTGA AGCATTGACC ATGCCACCAG TCACAGTTCC GACAACAGGA	6180
ATCACTTTTA GAATATTCCC GACTGCACTC TTACCTAAAC TAACCACTAC TAATTCTTTG	6240
GTTAAGCTTT TCCCTAAGCT TTCAGACCAA GACTGACCAA AAATTTTATG CAGCCGTGAC	6300
ATCATAGTTA ATTGAACAGG TACCAGTAAA AAAGCATCCG AAAACGGAAT TGGCGAAAC	6360
CCAATAATTG CTGCTGACAA TGCAGCAAAA TGAATCGTTT GATGACTTTT CTTTTTCACG	6420
TCAGGTGAAA CACCAGCCAA AAATTCATCA AATACTTTGT CGAGTTGTGT TTGACTGTTC	6480
GTCACTAATT GTTCTGCTTT ACCTTTAGAA GTTGCGAAAG CATTCATTAT AGCCGTAGAC	6540
GTTTTTTGTG GAAATTTTTT CAAGACATCT TGAAAAAAG AACGAAATTC TGGGTCTTCG	6600
ATTTCTTCAT TCAATTGCTT AGTTTCTTGT TCTGCTAATT CTTGTTTTTC AATCTCTTCA	6660
AAAAGTTCTT GTTCTTTTTT GTTTTATGT TTTTTCTTC TCATCTCATT AACTCCTTA	6720
TCTATCAATA TGTAGAAACA AAGTATAGCA TAGCCCCTGA TTGATTCTCT TTTAGATGAT	6780
TTGAAAAGCC AGAACTTATC TGTTTTTTAA ACAAAATAGAA ATGTTCCAGC TTTTCTCAAA	6840
AAAATCAGCT ATTTTGTTTG TTCTTCACCG CAATAAACCG TCAAGTGTTT AGGCAAAATC	6900
TTCAGCGTAA TTGGCAAAGC AGCTCCTTCA TCACCATCCA CATTAGTGGT TAATTCTTCT	6960
TGGTTTTCTA AGGAGATTGT TCCTTCTTTA AAGGTTAGGT AAATAAATT ATCAGTGGAC	7020
TGGTCGACGC CTTTTAACAA ATCTGGGACT GCTTTTACTG CATCCCATAA TGATTGATcT	7080
TTTAAATAAA CTAAATGAAG TTTGCCATCA TCCACTTGGG CCTCTGGTAA TAATGTTTCA	7140
AACCCACCGA TTGAATTCGT TAAGCCAACT AAAACGGTAC TACTTTCAAT GGTTTGTCT	7200
TTTTGATCCA AGCTTAAATG AAAAGGATAC GTTTGAGCAT TCGCTAAATG TTTGGCACCT	7260

GAAATAAAAT AAGCGAGTTT TCCTAACTTC GTTTTTTGT CAACGTCCAC GTCGTTAATT	7320
GA CTCTGGAA TCGTCCCAAT TGCAACGACA TTCATAAAGT AGTCGTCATT GATTTTCCCC	7380
ACATCTAAAG CACTGGTTTT TTCTAAATCT AATTGTTGAA TCGCTTCTTC AGGGTCCATC	7440
GGTAAATTCA AGGCTCGCGC TAAATCGTTC ACCGTTCCCTA AAGGAAAAAA GCCAAATTTT	7500
GGGCGGTAAG CTTGTTCTGC CAACCCGCTA ATTCCTTCAT TAACGGTTCC ATCTCCCCC	7560
ATTACGAAAA CACTGTCGAA GTGACTTTC GCAGCTTCAC GGGCAAATTG TTCTGCATCG	7620
CCGCCTTTTT CTGTTTGTTC CACTACAAC TCATCAAATA ACTGCTTTAA TTTTCTTCT	7680
GCTAACGTTT CAAATTCCTT GGCTTTTTCA CCACCAGAAC TGGGGTTAAC AATTAAACT	7740
GCTTTTTTCA CTGGTACCAC TCCTCATGTA TATCTAAATC GTTTGATTCC TTACAATTAT	7800
TATAAAAGAT ATTTTGTATT CTGGCGAAAA ATATGCTGAC ACTTGCTTAA CAAACAAGCA	7860
TTCTTGCTCC TCACTACTTT TCAAAACGTT CTTTAAAGCA ATAGGCAAAA TAATAGAACA	7920
ATTTCCAAAA ATCTTGCCAA ACTTTTGAT CCATCCCCGT CAATTCATGA ATTCGTTTCA	7980
AACGGTAATT CAATGTATTT CGATGGATAT GTAATACTTG TGAGGTTTCA TTGATGTTTA	8040
AGTTATGATT GCGCAAAAA ATAACGTTT CATAGTATTC ATCATAAATC GTCTGGATGG	8100
TTTCAAAAA ATCCAATTCA ATATGAAAAG GAAAATTAAA AATTCCTGA ATGAAAAAT	8160
CTTCTGCTTG GTAAACTTA GCCTTTTCTA ATTGGAGAAA CGATAGAAGT AAATACGTGT	8220
TTTGTGCCTC TTGGACCATT TGTGAATCG TGGCTTTTTC CGAACTTATC ACAAGGTATT	8280
CTGTTACTGG CTTTGTATC TGAGGTGGTT CAGAAGTAAT CACCAGATAA CAATTTTCCC	8340
ATTCAAATAT TTCATGTCCA GGATAATAAC TGGATAATAA GCGCTTGTC TTGGTCAGTA	8400
CACAATGATT CTGGTCATCT AATTGAAGGT TATATAGTTC CAACGCTTCA AGTTTAAAGG	8460
CTTCTGAGTA AGAAGAATCA ACATGCATTA AGTGCTGGAT AAAGTATTT TTTTGTTC	8520
TCTTTTTTTC GCGTCTTGT TGTGTATTTA ATTCTTCTGT CAGTAAAAGA ACAATCGAGC	8580
GCACAAGCTT GGTAATTTG CGCACCTCGT CTGGTTCTCC TGAAATGCCT ACGACACCTA	8640
AAATGTGATT ATTCAAAATA ATCGGTTTCAT TGGTGCCTTT TTTCTCTAAA GAAGTATCTT	8700
TATAAATTTT AATATTTTTT CGTTCATTGA TGGCTCTTTC TGCACCACGA TGGCGCTCCC	8760
CAATTCGAGC AGAATCTCCA CTCGCAATAA TTTTCTCTCG TTCATCCATA ATATTGATGT	8820
TATAAGGAAT ATCTGCCATC AACTTTTCAA CAATTTTATT TGCTTGATT TTGTCTAATT	8880
TTAACATTGC TATCGCCAAC TTTCTAAAAA GCGAGTaGCG TTTGGGACAT AAGCCGCAAT	8940
GACTTATGTC CCAAACgCTT AATTCTTCAG ATCAACGACT GAACATTTTT ATTTAGCTT	9000
TCATTACTGA TGATACCATG GTTAGTTCTT AGTCACAAGG CACCAGCTGT AATGTACGAG	9060
CAATATTTTC AACGGTTCTT TCTAAATTAA CTGGTCCATT TTTCAATGCT GCTTCCAATG	9120
GTTCGGCTTT GGCCAAAATT CAAAAATAG CCGTCATCCC TTCTTCATAG AGGACATCCA	9180
CATCTTTGCC GATATAGCCA GCACAAGCGA AAACCGGTTT TTGATACTTC TTGGCCAAAC	9240

GCAGAGCGCC ATAAGGGGCT TTGCCGAATT TTGTTTGAAA ATCCATACCG CCTTCACCAG	9300
TAAAGACATA ATCAGCGGCT TTTATTTTTT CTTCTAATTG ATTCATTCT ACGACAATAT	9360
CAATTCCTGG GCGTAACTGA GCTTTTGTGA ATGCTAGTAA CCCTGCACCC AAACCACCAG	9420
CAGCTCCTGC GCCTGGCATC ATCTCAACAT CAATGCCGAT CTCTTTTTTG ATGATCGCTG	9480
CATAATGGGC TAAGTTTTTA TCTAATTGAG CCACCATTTC TTCTGTCGCA CCTTTTTGCG	9540
GACCAAAGAC ATGAGACGCC CCTGTTGGCC CTGTTAGTGG GTTATTACACA TCACTTGCTA	9600
CTAGAACTTC GGTAGCAAAA ATGCGTTGAT CAAACTGTGT TAAATCGATC TGCGCCAGTT	9660
TATCTAATGC ACCACCGCCA CGTGTCAATT CTTGCCCTTC CTTGTCTAAC AAACGCGCAC	9720
CAAGTGCCTG AATCATACCT GCACCCCAT CATTGTGCAC GCTACCACCA ATGCCGATAA	9780
TGATTTTTTT TGCCCATGA TTGAGCGCAT GTTGAATCAT CTCTCCTGTA CCATAAGTAG	9840
AAGTGATCAA GGGATTGCGT TCTGCTGCTT TAACCAATTC AATTCCGTTG GCTTTCGCCA	9900
TTTCAATGAC CACCGTTTCT TGGTCACCTA ATAAGCCGTA ATAGGTTTCA ACGATTGAT	9960
CAGGAAGCGG TCCTGCCACG CGCACTGGTA CTCTTGTAAC TCCTGTTGCA TCCACTAAGG	10020
AATCCACCGT TCCTTCCCG CCGTCTGCCA TCGGGACATG TTGAATGTCT GCATCTTTAA	10080
ATACAGCTAA AATCCCTTTT TCCATTGCCA GACAGGCTTC TTTTGCACTC ATACTTTCTT	10140
TAAATGAGTC TGGAGCAAAA ACAAAATTTT TCGTCATCTT TTCGCGCCTA CTTTCCAAC	10200
AGTTTAAAAT AAAGCCATAT AAAATGGTTG CAACAATCGT CATTACCCCA CCGACACACA	10260
TTTCATAAGG AATCACTCGA ATTCTTTCTT TAATCGTCAT ATTCATACTT TGTGCCGTAA	10320
TATGGAAATA ATTGCCTTGA GGAACACTAT CAATCACAGT TGCACCAGTA TGCACCATGA	10380
CTGCTGCTGA AAGTGAAGAT GTGCCCATAT CTAAAATCGC TGGTCCAAAA GAACCTGTCG	10440
CTAAAATAAC CGCTGTGGAT GTTGAACCAA CTGCTAAGCC CATTAAAATA CCAGAAATCG	10500
GTGCCAAGAA GGTCCAGAA ATTCCCATAA CAGAAATTAA GTGGACAATT TGTGTTGATA	10560
AATCAGAAAC AGAAATTAAT CCAGCAATTG CACCTGCACC AATTAAAATT AAAACGGTAT	10620
CGGTTACTCT CTTCAAGCCA GAAGATGCAT AAAAGGCCAG TTGTTTTGTT TGCCCAATCG	10680
CAATTGTCCC AACAACTGAG GCGAACGGCA AAATATACAG AGCATCAATA TTCAAACTTT	10740
TGATGAAAGA AATATTTAAA ACGGCGCCAA TCGGATTAAG TAATAATAAG ACAATCGCTA	10800
ACACTGGCGC GATCAACGCC TTTTCAAAG GTAATAATTC TTTTGTGTTG TCCTCTACAA	10860
GATCTGTTTC TTTGACCATT GATCCTTTGT TTCTCATGAT ATTTGCAAGA ACAACAGCTG	10920
TAATTAAGCC AAATACGGCT GGAATAAAGC CAGCAATCAT AACGTCACCT AGAGGTAAAT	10980
CAAAGCCATT CGCCGCTGCA ATAGTATTTG GATTAGGCGA AATAATATTG CCCGCTTTGC	11040
CGCCACCTGA TAAGGCAATC AATAATGATA ACTTACTGAT ACCCGTACTT TTCCCAACAG	11100
ATAGAGCAAT CGGGGCAACA ATCAACACTG CAACAGGAAT GAATACGCC ACTGCTGTAA	11160
TTACCATTGT GGATAATGCC AAAGATAAAA TAGCTTTTTT CTCGCCAAGC TTTTGTACAA	11220

TGGCGTTGGC	AATTGATTCT	GGCGCACCAG	ATTCCATCAT	CACACCCGCC	AACATGCCGG	11280
CTGCTAAGAC	ACGGATTGCG	GTTCCCATAA	TACTTTGTGT	TCCGTTTAAT	AGAACAGCGA	11340
CTGTATCTGA	CAGGGAAGCG	CCTCCAATAA	GACAACCAAC	AATTGCACCA	AAAAACAAGG	11400
AATACACAGG	ATTAAATTTT	TTCAAAATTA	ATACTATCGC	TACAACTAAT	CCTAAAATTG	11460
CAGCAACCCA	AGATATTGTA	ATATCTGGCA	TTCAAAATCA	GCTCCTTAAA	TTGTTTACGC	11520
TTTCATTCTA	TCAAAGTCCT	TTCTTTTAA	AATTGTAAAT	TTAACAGGAT	TTATTGTTCA	11580
AATGCACAAA	AAAACCGTTC	TTAAAGAACA	GTTTTTTGAA	ATCTTTTCT	ATCTTTCTCA	11640
ATAATGACCA	CCAATCCAAC	GaACAAAAGA	ACACTCACAA	TTAACAATAC	TGCCGTTTAA	11700
TTATATCCAT	GAAATAGCKT	cGCCTCCACA	TGCGTAAAGT	ATTGAGGAAG	GAATTACACC	11760
AATGAATATA	GCGGCTATTA	AATTTTTTAA	TGGCAACTGC	ATAGTATCTG	CAGCAAAATT	11820
AATCACAGAA	GAAGGAATAA	TTGGAATCAT	GTAGCCCAAC	ATAATCCCAA	TTTTTGGATG	11880
CTTCATCCGT	CGAATGGACT	GGACCCAATG	ATTGGTCTCC	GTTTTTTTAT	CTAAAAATTT	11940
TATATGATGC	ATCAAAAAGA	TTGCAACCAA	ATTTCCCATA	CTATCCCTA	AAACGTTAAT	12000
GAGTGTGCCG	CTGAAGGGAC	CAAAACTCAA	TCCTGCAATC	ACACCAATTA	CTGAGGTCGG	12060
AATCCCTGGT	ACTGCACACA	TGATCGCGGT	CAAAAAAACC	AACAATAGCG	CTGCTGAAAA	12120
ACCATGTGAA	CGAATTCGCT	CTAATAATTG	TGTCTTACTG	GCTTTAGGAT	TCAAAAATAA	12180
TTGAATATCT	GTTTGATATT	CTAGATACAA	ACGGTAACCT	ACCAGCAGGA	TTAAAATAAG	12240
CCCGCTACTG	ATTAGTAGTA	TTCGGATTGT	TTTCTCTTTC	ATTTTCTTTG	CTCCTTTTAT	12300
GCTTTATACA	AAGTGTAGCA	QAGGATTCTT	TGTGCCTCAA	TTAGAAGCTT	GTATGAATAA	12360
AAGCCGCCCC	TAAAAGTTAG	ACTAAAAATC	TAACTTTTAG	GGGCGATTCTG	AAAGTTACGA	12420
AGATACTTTT	TTTATTTCCG	ATACATCTTA	AACTCCAAGA	ACAGATCGTT	ATAAATTCCT	12480
AAGTATTCTT	GACCTAAGTC	TTGGTAAACT	TCTAAATGTT	TGATAGTTTC	ATCATCTGGA	12540
TAAAATTGTT	TATCTTCGGC	AACTTCTTTT	GGTAATAGTT	TTTLAGCTTC	TTTATTTGGT	12600
GTGGAATAAC	CAATATATTC	TGCATTTTGT	GCCGCATTTT	CTGGTCGTAA	CATAAAGTTC	12660
ATAAATGCAT	AGGCACCCTC	TTTATTTTTG	GCTGTCTTAG	GCATCACAAT	GTTATCAAAC	12720
CAGAGATTAG	AGCCTTCACT	GGGAATCACA	TAATGTAGAT	GTTTATTGTT	TTCTAGCATT	12780
TCAGCAGCTT	CACCAGAGAA	AGTTACAGCA	ACTGCACTTT	CTTCATTAGc	CATGTACATT	12840
TTGATTTTCAT	CGGCAACAAT	TGCTTTGACA	TTGTTTCGTTA	ATCGGTTTAA	CTTATCGGTA	12900
GCCTGACGTA	ATTGTTGGTC	GTTTTTACTG	TTTAACGAAT	AGCCTAAACT	GTTCAAAGAT	12960
AATCCTAACA	CTTCGCGAGC	GCCATCAATC	AGCATGACAT	TATTTTTTAA	TTCCGGGCGC	13020
CATAAATCAT	CCCAATGTTG	GATCTGACGG	CCGTCAATAA	ATTTATCATT	ATAAATAATC	13080
CCCAACGTGC	CCCAGAAGTA	CGGAACGGAA	AACTTATTTT	TGGGATCAAA	GGATTGATCT	13140
AAAAAGCGTG	CATCAATGTT	TTCTAAGCCT	TTTAATTTGC	TGTGATCAAG	TGGTAAAAGC	13200



ATCTTCGCTT	TCATCATTTT	TTGAATCATA	TATTCAGAAG	GAATGGCAAT	ATCATAGGCT	13260
GTGCCACCTT	GCTGAATTTT	TGTATACATA	GCTTCATTAG	AATCAAAGGT	TTCGTAATTG	13320
ACTTTATAGC	CTGTTTCTTT	TTCAAATTTT	TTAATCAAGG	CCGGATCTAT	ATAGTCCCCC	13380
CAATTGTAAA	TGGTCAAGGT	ATCTGCTCCT	GCCATGCCAC	TTGCTTTCTC	CAATTGGCGC	13440
ACGCCAAAAA	ACAAGATGAC	AATAATAGCG	ATAATTCCAA	TAAAAAGTGA	CTGTAATTTT	13500
TTCATTTTAA	CTTGGCCACC	TCACTTTGTT	CCCACGCGAT	TTTCTTTAAA	CGTTTGGCG	13560
TATTGTCTTT	ACTGATAAAG	TAATAACCGA	CAACTAGAAT	CATTGAGAAT	AGGAATACTA	13620
ACGCACTTAA	GGCATTGATT	TCTAAACTAA	TTCCCTGACG	TGCTCGTGAA	TAAATTTCAA	13680
CAGATAATGT	TGAAAAACCA	TTGCCTGTCA	CAAAGAATGT	TACGGCAAAA	TCGTCCAACG	13740
AATAAGTAAA	GGCCATAAAG	TAGCCAGCAA	TAATTCCAGG	CGCTAAAAAA	GGCAAAATAA	13800
TATTTTTAAT	CACTTGAACA	TTATTGGCCC	CTAAATCCCG	TGCCGCATCT	ACCATAGAAT	13860
CGTTCATCTC	TTGTAGTTTT	GGTAAAACCA	TTAAAACAAC	AATTGGAATA	CTAAAGGCAA	13920
TATGCGACAA	TAATACACTT	GTAAAACCTA	AACTGATAAA	CCCTACTGCC	GTGAAGAAAA	13980
TTAAGAAACT	GGCTCCAATA	ATAACATCTG	GTGAAACGAG	TAAATGTTA	TTAAAGCTAA	14040
GCAACGTGTT	CCGGTTTTTA	CGACGTTTTG	TATAATAAAT	CCCCATCGCA	CCAAATGTTT	14100
CAATGATGGT	TGCTATTAAA	GCCGATAAGA	AAGCTAGCAA	AAACGTATTC	AGAACAATGG	14160
TTAACAAGCG	TGTGTCTTCG	AAAACGTTCC	GATAATTATC	TAACGTAAAA	CCAGTAAAcT	14220
GATTCATCGT	TTGCGTGTCT	TTAAACGAAT	AAAAAAAtCAA	ATAAAAGATT	GGTGCGTACA	14280
GTAGCGCAAA	GACAACGAAT	AAATACACAT	AAGACCATTT	AAACTTTTTT	TTTTGCATTA	14340
TTTGCCACGC	CCTTTCTTCT	TCTCACCCTG	CAGCAACATC	ACAATAAACA	TGGCAACAAT	14400
TAAATCACA	CCTATTGTTG	AACCCATGCC	CCAGTTTTGC	GTTACTAAGA	AGTGTTCCTC	14460
AATCGCCGTT	CCCAACGTAA	TCACCCGATT	GCCGCCAATT	AAACGCGTTA	ACATAAAGAG	14520
CGAGAGCGAA	GGAATAAAGA	CAGCTTGAAC	GCCACTTTTC	ACACCGTTTA	GGGAGAGAGG	14580
AAAAATGACA	CGGCGGAACG	TCTCCGCACT	GCTAGCTCCT	AAGTCCCGAC	TAGCACTAAT	14640
CAATGAAGGA	TTCATCTCTT	CTAAGGCATT	AAAAATCGGC	ATAATCATAA	ATGGAATTTT	14700
AATATACGTT	GCGACAAATA	AAAAACTGAA	ATCTGTAAAC	AGAATCTGGT	GCGTACCCAA	14760
ACCCATAAAA	CTCATAAACT	GATTGATACT	CCCATGAATA	CTGAAAATCC	CAATAAAGGC	14820
ATAAGCTTTT	AATAATAAAT	TCACCCACGT	TGGTAAAATA	ACAAGCATT	ACCACAACCTG	14880
TTTATGCTTG	AGCTTTGTCA	AAAAATAAGC	CGTAGGATAA	CTTACCAATA	ACGTAAAAAT	14940
GGTAATCAGA	AAAGCATACC	AACTGAATT	AATGGTCATA	CTTAAATACG	TTCCTGATGT	15000
AAAATACGTT	TGGTAATTCT	CCAAGGTAAA	CTGGCCATT	ATATCGAAAA	ATGATTGATA	15060
AATAATCATT	AGCACTGGCG	CAATGACGAA	TAAAAAGTAAC	CACATTACAT	AGGGAATCGA	15120
ATAAATTCGA	CGCATTGTTT	TCATTTTATT	TTCCCTTTT	GTTTACTCGT	CGTAACTTTC	15180

TAAACGGGCA TCAAAGTCTT CTTCTGATTC ATTGAAGCGC ATCACGTGAA TATCTTCTGG	15240
CTCAAAAGAG AGCCCTACTT GAGCGCCTTC TTTGGCTTTT CGAGTGAAT GCACCATCCA	15300
TTCGTTTCCT TGTTTCATCAA AACAAATAAT TTCATAATGA ACGCCGCGGA ATAAGTGGCT	15360
GTCCACCGTT ACGGTTAGCT TGCCTTTTTT AAGCGTGGTA ATCGTTAAGT CTTCTGGACG	15420
TAAACGACT TCTACAGTT CATTTTTCCG CATGCCGCCA TCGACACATT CAAACTTTTT	15480
GCCACAAAT GAACTAAAT TATCTCCAG CATCACGCCA TCCACAATAT TACTTTCCCC	15540
CACAAAGTCA GCAACAAAGT GATTAATCGG TTCATCATAA ATATCGACTG GTGTGmCACT	15600
TTGCACGATA TGACCTTTAT TCATGACAAA AATTCATCA CTCATCGCTA ATGCTTCTTC	15660
TTGATCaTGG GTCACAAAAA TAAAAGTAAT TCCTAAACGT TGTGTAAAT CACGTAGTTC	15720
ATATTGCATC GCTGTGCGTA ATTCAAATC TAAAGCGGAT AACGGCTCAT CTAAAAGCAA	15780
TACTTTTGGC TCATTGACAA TCGCTCGTGC AATTGCGACC CGCTGACGTT GGCCCCcAGa	15840
CATTTCACTA ATTTCaCGTG TTCGTAGCC CGGTAATTGA ACCAAACGTA ATGCCTCTTT	15900
CACTTTTTTT TCAATTTCTG CTTTGATAG TTTTTGATT TTTAAACCAA ACGCAACATT	15960
ATCAAAAACA TTCATGTGAG GAAACAAAGC ATAATCCTGA AACACGGTAT TGAATTGGCG	16020
TTTATTGGCA GGTACATCAT TTATTCTTTT ACCATCAAAA TAAACATCGC CAGAAGTCAC	16080
GTCATTGAAT CCCGCAATAA TACGCAAAAT CGTGTGTTTG CCACATCCTG AAGGTCCTAA	16140
TAACGTATAA AATTTTCCTT CTTCAATTC AAAGCTGACA TTTTAAATA CTGGTTCATC	16200
GTCAAATTGT TTGACAACAT TTTCAAATGA AATCACATGG TTTCTCACTG GATAAACTCC	16260
TCCATTTCTG CTTTATAAA TACGATTCTG TCGCCACAAT TAACGCACGA CTTTTTCCTC	16320
GACCATTGTT GCTAATTTGG TGAGGTTCTG TTGCTTGGTA ATACATGGTT TGACCTTTTT	16380
TAGCAATATA CGTGGCTTCC CCTAATGTCA AACGAATGGA GCCATCAATC ACATAAATAA	16440
ATGTTTCTGA TAGAGAAGGT TCAAACGTTT TGTATTCGCC ATTTTATCA AAAGTAATAA	16500
TGACAGGTTC CaTTTCTTTT TCATTCGAGG CTGGAATTAA CCATTTTAGT TCATAGCCAT	16560
GTTCTTCATC ATAATAGAGG GACTTTCTT CTTCAATTGA GACAATTTTC TGTCAAGCG	16620
GTTGCTGACT AAAAACTGC TCTGGCGTCA CGCCTAGTAC TTCTAAAATT GAGAAAAATG	16680
TTTCCATTGA TGGTGAGCTT AAATCCCGCT CCAATTGGGA AATATAGCCT TTGGATAAAT	16740
CCGTTCTTTC GCCTAATTCT TCTTGTGTTA AATTTTCTG AATGCGTAA TTTCTCAACT	16800
TCTACCAAT TTCCATGGCA CACCTCGTTT CTATTTTCAG CTAATCTTTT CAACAAAAGT	16860
TTTGAATAC	16870

## (2) INFORMATION FOR SEQ ID NO: 99:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3905 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

GAATTATTTT AATTACCTGT TGA	CTATCAA GATGAACGAT TAACA	ACTGT CCAAGCAGAG	60
CGTTTCTTGG TGGAGCAAGC	AGATGCTTCA AGAGCCAAAA	GAAAAAAGT GATTGATAAA	120
CTAGCGGCAG TAATGATTCT	ACAAAATTAT TTAGATGCAC	ATAGCCGATA ATTATCGAGC	180
TTTGAATACA TCAnGGGAGA	GAAGAAGAAT GACAGAACAT	AACCACGATC ATGATCACGA	240
GGACATGAAC ACATTACTTT	AGTAGACGAA CAAGGAAATG	AAACATTATA CGAGATCTTA	300
CTAACTGTTG ACGGTCAAGA	AGAATTCCGT AAaAATTATG	TTTTATTATA TCCAGCAGGA	360
ATTCTTGAGG ATGAAGATGT	TGAATTGCAA GCTTACTCCT	ATGTGGAAAA CGCAGAAGGT	420
ACAGAAGGCG ATTTACAGCA	AATCGAAACA GACGCTGAAT	GGGATATGAT TGAAGAAGTT	480
TTCAATACAT TCATGGCTGA	AGAAGAAGAG TAAAACTGTA	ACGAGAGAAG CGCTACTTCT	540
CCTATTGAGA AGTAGCTTTC	TTAATAATAA TTGTTATAAT	GTACATGTTG TTTCTATCAT	600
AAGATACTAT CACTATCATT	GAGTGTGATG GATAAAGAAG	GACAGCATTG TCTAGTTTTA	660
CAAGATGTTG TAAAAACGTA	ACATCCTTGT AATGACGTTA	ATTTAGTAAC TATCTAGCAT	720
AATTTATAAT TTGTGTTATA	CTAAGTATGA GGTTTTAGTA	ATACTGTTAA TTTkAkAaTA	780
TTAAATAAAA AAATGGCTCC	ACCTACTGGC ATAGGTAGAG	CCGTCAGAAA TTCAGTTAAA	840
CAAACACGTA ATTATTGGTG	TGGTGCAATTG CCACGAACTG	AAATTCCAGT TAATAAAGCA	900
AGTAAATCAA ACATATATTT	CACCTCCTTT ATTATGATAG	CTTCATTATA ACATTATCTG	960
TAAAATGATG GTATGTTTAA	TACAAAAAAA ATAGAAAGAT	TGGAGTTTGT ACTGTGAGAG	1020
TAGCGGGAGA AAAAATCAAA	ACAGCGCGTA AAAAGAAAAA	ATTATCGCmA GCTGAATTGG	1080
CAAAAGGAaT TTGTACACAA	GCmACAATTA GTAACATTGA	AAACAAAAAT GTGTGTGACA	1140
GTTTAGACAT TTTCTCATCT	GTATGTCTAC GTTTAGACTT	ACAAGTAGAA GAATGTATCG	1200
AAGGTTCTAG TGAGAAAAAA	CTCGAAAGCT TATTAAATAA	AGTAGAAGAA TTATGCTTTT	1260
ACTTTAAGCA CGATGAAGCC	TATGATTTGT TGAAAGACTA	TCCAGATGAT ATCGAAAGTT	1320
CAAACAGAAT TTTAGAAACA	AAATTTTCT ATTATAAAGG	AATTACGAGT TTATTAGGTA	1380
AGAAAAATAA TAGCGAACAT	TGTTCTATTT ACATCGTGGT	TCTGAAATTA GTCGTGATAT	1440
TAATATTTAT AATATTtTGA	GTATGAATGC GATCGGTATT	TTATATGAAT TAGAAGATGA	1500
TATTGAAAAA GCAAAAGTTT	ATTACGATAA ATCTTTACAA	TTGCTAAGCG AATTCAAATT	1560
AGATTATCCA TTAGAGCAAT	GCCGTATTTA TTATAATACC	GCAAAATTCT ATTCATTAAT	1620
TAAAGACTAT GCAAAGAGTA	TTGAATTAAG TGATAAAGGA	ATCGAAATTA ATCGGACGCA	1680
CAGTTCTATT TACTCTTTAG	ATTGCTTATT ATACGAAAAA	GCGTTTAATA AACAAATGCT	1740
AGGACTAGAT GCTGTTGAAG	ATTATCGTAT TGCGTACTAC	TTTACACGTT TCTTTGAAAA	1800
TAAAAAATTA CTTACTTATA	TTGAAAAAGA TATGCAAGAG	TTTAATATCA GTTTTAAATA	1860

ATAAATTATT	GTTTCAGATT	TtAAAAAGAA	CAACCTAACT	CAAGAGGAGT	TAGGTTGTTC	1920
TTTTTCTCT	TTTAAATTTG	TTTCTTTGAC	AGCGTTACGG	ACGCTTTCGG	CGACGACTTT	1980
TGCGACTCCT	TCTTGAATG	GATCGGGAAT	AATATTAGTC	GGCGTGAGTT	CATTATCGGG	2040
AATCAATTTT	GCAATACCTT	TGGCAGCGGC	AATTTGCATT	TCAATCGTAA	TCTTTTTCGC	2100
CCGAGCATCT	AAGGCACCAC	GGAAGATACC	AGGGAAGGCT	AACACATTAT	TAATTTGATT	2160
CGGAAAGTCA	CTACGACCAG	TTCCGACAAT	ATAAGCGCCT	GCCGCTAAAG	CTTCATCAGG	2220
AAAAATTTCA	GGAAGTGGT	TAGCCATTGC	AAAAATAACT	GGCTGCTCAT	TCATCTGTTG	2280
AATCCACTCA	GGTTTTAAAA	CACCTGGTGC	CGAAACACCA	ACAAAGACAT	CAGCGCCTTC	2340
AAGAGCAGTA	GCTAAATCAC	CGGTACGgTG	TTcCCGATTG	GTTAGcTTCG	CAATTTCTGC	2400
GTGGTGTGGT	GGCAGAGCTG	TATCTGTTTC	GCTCAGAAcTC	cCTGCACGGT	CAACAATGAT	2460
AATATGTTTT	ACGCCAGCAG	CTAAAAATTT	TCTTGTGAtC	GAAAGACCAG	CGGAGCCACC	2520
GCCGTTGATT	ACCACATGAA	TGTCCTCTAT	TTTCTTGTG	ATTAATTTTA	AGCTGTTGTA	2580
CAAAGCCGCT	AGAACAACAA	TCGCTGTCCC	GTGTTGATCA	TCATGGAAAA	CGGGAATATC	2640
TAGTTCATCG	ATTAAACGCT	GTTCAATTTT	AAAACAACGA	GGAGCGCTGA	TATCTTCCAA	2700
GTTAATCCCC	CCAAAGGTTG	GAGCCAAAAA	TTTCACCGTT	TGAATAATTT	CTTCGGTATC	2760
TTGTGTATCT	AAAACAAGGG	GAATCGAATC	CACACCAGCA	AAGCGTTTAA	ACAATGCAGC	2820
TTTTCTTCC	ATGACAGGCA	TCGCTGCTTC	AGGACCAATG	TTACCTAAAC	CTAAAACTGC	2880
TGAACCATCA	CTAATAACGG	CGACGGTGTT	TTTCTTTGTC	GTTAATTCGT	AAGCGCGTTC	2940
TTTTTTTTCA	TGTATGGCAG	AAGAAAACAG	CGCAACACCT	GGTGTGTAGG	CAATTCCTAA	3000
ATCTGCTTTT	GTTTCAATCG	GAACTTTAGG	GCAGACTTCT	AATTTCCCCC	CAAATTTTTT	3060
TGCTTGCTCA	ATCGCTAATT	CTTTTACATC	TTTTGTGAC	ATATCATTTT	CCTCCTAAAT	3120
AATTCACCTA	TAGTTAAAAA	AGCATTCTAA	AAATAGCAGT	CATTGTTATT	ACTGTAATGG	3180
CACCGCCTAA	ACGTGTCGCA	ACTTGTGCAA	ACGGCATTAA	GTTCAATCGA	TTGGCTGTAC	3240
TTAAATAGC	CACATCGCCA	GTCCCACCCA	TACCACTCTG	GCAAGCGGAG	ACAATGGCAG	3300
CCTCAACGGG	ATTCATATTC	ATGAAGCGTG	AAACGAAAAA	TCCTGTAGCA	ATAAygGTGA	3360
AAACGACACT	AATTACCACA	ACAAAATATT	GCCAGCTAAG	CATTCCTACG	ACATCTTTTA	3420
AAGGAATATA	TAACATACyT	AAGCCGACCA	TTAACGGGAA	CGTAAAATTG	CCAGAGATAA	3480
ATTTATAAAG	CTGTTTTGAT	CCGCGTTGGG	TTTCTTTCGG	TACTACGTTT	AAGTATTTTA	3540
AAAAGGCTGC	GACAACAATC	ATCAAAACAG	GACCAGGAAA	GCCTGTTAAG	TGTTGCAATA	3600
AACCGCCTGT	AATAAATAAG	GTACACGCAA	TTAAAACACC	AGCGCCCAT	AATTTTACAT	3660
CAATTGGTGC	TTGTCTTCT	TCTAAAGCAT	CACTTAAATC	ATCTGAATTG	GTAATTTTAA	3720
TCAACTGACC	TTGTCCAGAA	AGTTCAGGGC	GTTTTTCTCC	TAAACGACTA	AGTAAACCAG	3780
AACACATAAT	AGCAAAAAAA	TTACCAATAA	TGGTTGCGGG	AATTAATTGC	CCCACTAATT	3840

GTTCACTTGG CAATCCAGTA ATGGCCACTA TAACCAAGGG ATAAAGGAAG AATGCCTTCA 3900  
CCGAT 3905

## (2) INFORMATION FOR SEQ ID NO: 100:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32768 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

TAAAAAGCCC CTTAGAATCA TAGGTnTTTT TGATGGTTGT AATTGTTGAT GTTTTCTTnA 60  
AGTTGTAAAG ATTTGATTTA ATAGTGATCG TGTCACCTAA TTGATAATCT TCACCTAATT 120  
TAAACaGTTT AGAGCTTGTG GGAACCTCAC CACTTAGTGT TAAAATTCGT TTACGTTCTG 180  
AAAGTTTGTT ATTCCCTCTA TTTTTCaAGA CTTCGAGGTA TTGTGAATCT GTCAAAGTGA 240  
CCTCGTCAGT TGTCTTTTGT AAGTCTCGAG CATCTACGTA TAACTCTTTT CTTTGTAGAC 300  
CTGTTTTCTC GTTACCCACC ACTACACTTT TACGTTCCGA GCC TTCACCT TCACCGAAAA 360  
CAATAGCGGT TGAAGATTCA TCAAAGTTAT TGTTCTGATA CCCCCTTTT GTTAGATTTT 420  
CGTATTCGTC TGAAAACTCA ATCGTTCTCG AAACATCTCG ACCTTTAAAA ATAGATAACG 480  
TATTACCAGG AACGCCTAAT TTAGTAGCTG TCTCTCTTAT CCCAAAATCG TAGGATGTGC 540  
AAAGAGTTTC CACTTCTTCG GCTACGACAC CATAACTATT CTGATACTGG ATAGACGTAT 600  
TCCCTAAATT AGCTCGGGAA TCCAATTTTA AATATTGAAT TTTTCTTTTT ACATCCGAAG 660  
GgTTTACTAC TTCATTATTA AGATGTTCCC AAACAATTTG CTCTGGCGGT GCCGTTTTAT 720  
TATAAATTCG GTAAACAATA CGATCTAATG ATTTCCAAG AAGTGATTTC CCCGACACCT 780  
TTATTTGTGC TGTTGCCTGA TCGTCTACAA CTACGGCATC AACATAGAAA TAGAGCCCTT 840  
CCATATAAAT AACAGTGTCT GGAACAAATA TTTGTATATT TTCAGGTGTC AAACCAACAA 900  
ACAGCTCAAA CGTTGAATAG GTATAGTAAT TTAATTTAAC AGTTAACTT TTAAACCCAT 960  
CAAATACTTT TTCGGAAATA AATTTCCATT GATTATCTCG ATCCTGAGTG AAAATTTCTA 1020  
ACTCCATTAC ACACCACCTA CCAACGGCCG GAAGTCAAGT TGTACAATTA CGTTACCAAT 1080  
ACCAGTAGTG GCTTTCACCT GAAAATAGTT ATCACCCCTT TCAAGCTGCA GAAAGGTAGA 1140  
ATTAGGGTCT CGCAATGGCA TAGCGTTAGA TTCAACACCG TTTGGGTCCG TCATGATTGC 1200  
TTCTTTCTGT CCACGAGTAG TTACCAGTTT AACTTAGTA CCTGCTCAA ATGAACCTTT 1260  
AAATCGGAAA AATTCCTGAG TAATTACGTT ATACACTTCT GGATCGGTAG CTTCTGCACC 1320  
TAAAGAAAAA TAGAACGTGC CTCCCACTGA TACATCACCA TCATTGGTTA CAGGAACAAT 1380  
CTCACCCTT TTTAGTGTAG CAAACTCATA CCCTTTCGTT ATTTCTAACG GCCACATGTG 1440  
TTTTTTTGTG GCAACGGCTA GGGGAATTAA CGTATTATAT TTAATCATAT CAGACCAATA 1500  
CGGATCTAAA GCTAGAAAGG TACACGTAAA AGCCTGTGTA ATATTTTTAG CTGGATCTGG 1560

CAGGTCGGGA	GCTTTTACTA	CTAACACATC	AATTTGAAAC	TCACGATCAA	ATACTCTGTA	1620
AATCAACGTA	CCTGGTGTTT	TGGGATTTAA	CGTTTTTATC	ATTTTCATGCT	GTAAATTGAA	1680
AAGATCCTTT	GTTCCTCTAG	CAATTATTTT	ACCTTTTATC	GTTAAATCGC	GCTTATCTAA	1740
ACGCTCCGAA	ACTTTAATAA	CACCATCTAG	TCCATATTGT	TCTTCTGTTA	TGATCTTATT	1800
TTCAACAGCT	CCGAAGCCTG	TTTTACTTTT	AACGGTAAAA	GGTGGTTGAA	CACCGAAACT	1860
AATCAAAGAA	CCTTCCGAGT	TCTTGATAT	AAGTTCGTAC	ATCCAACCTA	CCTCCTATCT	1920
TCTCAGTGTT	GCAAGGTCTT	GCAACTTGTA	TTTTGTTTCT	CTTGCAATTT	CTCTAGGCGT	1980
TAAAGGCTCG	GGACTTGTA	TATATTGTGT	AACATTGATA	TCTCCGTCTT	TTTGTTTCAGC	2040
CATTACACCT	CTCACTGCAG	TTCAACATA	TCCTAACAGC	GTGTCAATCG	GTGCGACAGC	2100
CTCAGCCCCT	GCTTCTCCAC	CTATCATTGC	GTTATTTCCG	TTCATGCCAA	ATATGGTAGG	2160
TTGCGTCATG	ATACCGCCGT	CTTTGTACCA	TTCTATGCCC	AAACTTGGTA	TTTTACCTTT	2220
CAATAAATCG	CCCGCAGACC	AACCGGCCGG	ACTAATAGAA	AAGTGAGGCA	GCGGAATTTT	2280
CGGCCAGCTA	ATATTAAAGT	TAAAGAAACC	TTTAATCGCA	TCTACTACAT	TTTTCACTAG	2340
ATTTTTAGCC	GCATTCATTG	GCCCATCAAT	AGCATTTTTT	ACACCATTAA	AGATATTAGA	2400
AACTGTTCCCT	GTAATACCAC	CCCATACATT	TGAAATAGTA	TTTTTCACTG	CGTTAACTAC	2460
ATTACTAATA	GTATTTCTTA	CACCATCCCA	AATAGATGAA	ACGGTATTTT	TCATACCGTT	2520
AAATAAATTC	TTAACGGTGT	TCAcCGTTGC	ACTAACGATA	TTAGACACGG	TCGATTTAAT	2580
ACTATTCCAC	ACCGTGGATG	CTGTATTTCT	AATTGCGTTA	AAAATGTTTG	TAATTGAATT	2640
TTTTAAAGCA	TTAAAAATAT	TTATCACTGT	GTTTTTAACA	GAAGTAACGA	TATTGGAAAT	2700
AGTCGTTTTA	ATACTATTCC	ATACACTAGA	AGCCGTGTTT	CTAATGGCAT	TAAAAATGTT	2760
TGTTATCGTG	TTTTTCATAG	TATTAAATAT	CGGTGTAACA	AAGTTAAGGA	TAGTCTGCAC	2820
AACATTAGAG	ATTGTTGTTT	TAATAGCATT	CCAAACGTTA	GTTGCTGTGT	TTTTAATTGC	2880
GTAAAGACA	TTAGTAATCG	TCGTTTTAAT	AGCATTAAAT	ATAGGAGTGA	CAAAATTTTT	2940
AATAGCCGTA	GAAACTGAAG	TAAGTGTATT	TTTAATTCCG	TTCCAAGCGA	TTGTTACAGA	3000
GTAAACGACA	GCATCCATA	TCTGTTGAAA	GAAAGTTTTA	AACGCATCAA	AAATACCCTT	3060
AATAACTCCA	ATGAAACCGT	TAATTATTCC	TGTGACGGTA	TCACCTATTT	TTTCCCATAC	3120
ACCGGTAGCA	ATTGCTTTTA	TGAGATCCCA	TAGATATTGA	AAATACGCTT	TAAAACCATC	3180
AAAAATACCT	TTCCAGATAT	TAATTGTATT	AACGATTATC	GCCCCTACTA	CAGCGTATAT	3240
TACATTCCAC	ACTGTTTGTA	GAAAATTCTT	AATGCCGTTA	AAAACGTTTT	CTACTGTTTT	3300
TTTGAAACCA	TCAAATGTTT	TCTTGCCAGC	ATCCGTGAAT	TTCTTCCAAA	TTTCTGAGAC	3360
TTTAGCGGTA	ATATCATTCC	ACGTTTTTTAC	TAAGCTACCT	TTTAACCCAC	TAGCTGTTTT	3420
AACGATTGAA	TCCCACGTAT	TTTTTATAAA	ATCGGCTATA	CCATTAAATA	TCTCTACTGC	3480
TTTATCTTTA	ACGGAGTTCC	ATGTATCAAC	TAGAACTTA	GTAAAAGATT	TCCAAATCTT	3540

TTGTCCTGTT TCAGTTTGGG TGAAAAAATA AATTAACGCC GCTACTAACC CCACGATTGC	3600
CGAAATAATC AACACAAAAG GGTGGCGTT TAAcACTGCG TTAAAAGCTG TTTGAATTGC	3660
AGTATAGGCT TTTGTGACAC CTGAAGCTAT TTTAGTAGCT GTTAATACAC CTTTCATGAT	3720
AATTTTATAG GTTGCTAAGG CGGCTCCTAC TCCAACATA ATAGCTTTTA AAGAACTAAA	3780
AATAGCTTTA TTTTGTAGTTA TAGCCTTTGT AAAATTAGAA AAAGCAGGGA TGACATTATT	3840
TACAATAACC TCACTCACAG CTCCTGATAT GTCCCAAATC ACCTCTACGA TTTTTTCAAT	3900
CGCAGGTAAT ACTACTGTAT TTATTGTTTT AAATGCAGAT TTAAAAATTT CTATCGCTTT	3960
AGGTACTACA GCTGTGAAAA TATCGGCAAA AGTAATTGAC CATGTACTAA ATACTCCTGC	4020
CAATTTTCGTA AATATCATAG ACACTTGACT GCCCGTATCC TCAAAAATCC CTTTTATTTT	4080
AGGCAGTTCT TTTTGTACAA TCTGAGCTAC ATTTTTTATA GCATCACTAA TCCACTTAAA	4140
AGCCCACTGA AGATTTCATCA CAAAACCTGA ATTTTGAATG CCGTTAAATG CATCCATAGC	4200
CGATTTTTGG AATCCCCCAA AGTCTGGTAT AACTGCAGCT ACCTTATCAC CTAAGCCACT	4260
TAAGCCTTTC GTAATGCTTT GAATAATCGA AACGACCGTT CCTAAAATAG GAGTGCCTAT	4320
TTTTGCTAAA AAGTCCTGCC ACGCTTGTTT TAAGTTACCC ATAACGTTTT CATATCCGTC	4380
GGCTTCGCGA GCAGCCTGTC CTACAGCTCC GGCTTGTTTA AGCATGTTGC TAGCATAATC	4440
TAAACGAGTA GCTTGTTTCG TCGCTTCATC TAAATTTTGC CAATCTTTTG TAGAACCTAC	4500
TACTCCCTGA GAAATAGCAA ACTGCGCCAT TTGAGTATCA TTTGCAAAAA TCCCTATCGC	4560
TTACCCGGCT TCGTAGTTTC CTTTGTAGAAA CGAAGTAAGA GAACTTGAAG CATCTTCCAG	4620
TGAAACATCA TAGAAAGCTG CAGCATCTGC CGCTAAGGTA GTTGCTTGCT CTGATTTCTT	4680
CATTGCGGCT TCTGTATCTA GCCCTAAGCC TTTAAACATT GAAGTAATTC GAGAAAATGT	4740
CGGCTTTATA CGGTTAGGTA GGATATTCAT TTTTTTACCC ATGCCATCTA CCATTTTTTG	4800
TGCTAATGGT TCTAACTTAC CAAAACCTG GGTAAATTGT GCATTCAAAG CTTGTGCACT	4860
TGCCGCGGCT TCTACGCTTA GCTTCCCAA ATTAACGATT TTTTCTACGG CAAAAGCAGC	4920
ACCAATTGCT AGCGCTGCTT TTTTAAAGAA CCCCATAAT TTACTTGTTG TTGATTGTCC	4980
CTTAGATTCC GTCTCGTCCA AAGCCTTATT GGCTTCACCA TTATCGACAC CGATTGTACC	5040
AAAAAGCTTG AAAATTTTCGC CCATTAAGAC ACCTCACCTT CTTTTCGAGG TTTAACAAAT	5100
TGACTAGCAA AATCTAATGC TTTTGTGTGT TCTTCTTTTG TGGTAGATTT CGCTTTATTT	5160
TTACGTAAAG GTCTGTATTT TTGTTGGTCT TTAAATTCTT TTAAAGACTG CGTCATCTGT	5220
GTGTGTAACC ACTGATTGTA GATTTTTTCT TCTTGTTCCA CTTCTTGAAC ATAAAAAATA	5280
AATTCCATTA AGTCACACAA ACGATAGGTA CGAAGTAACG ATATCGGATC TCCGTACCTT	5340
TTGAACAACA AATCTTTAAT TCGGTGTTTCG CCGTCTTTTA TTTCAATAAG CTGGCGATAG	5400
AAGAGAAAAA ATCAGCTAGT TCTGGTTTCT TAAAGAAATC GACTAATAAA GTCGTATACT	5460
CCTTTAATCC AAGAGCTGAA ATTTCTTGTA TAGAAACGCC TGTAACTCT GCTAAAAGTC	5520

CGTTAATAAC	ACTCTTAATA	TCTTTTAAGT	TAAGCATGAC	TTTTTGCAAC	ATCGctGCGC	5580
CCATTTCAAT	GCCACGCTTA	GTAGCTTCTT	TGTCAGCTTT	TTCTTTGGCT	GCATCAATTT	5640
TTTTCTGCTC	TGCTTTTGTT	GGCTCTTTTG	ATTTGTGGTC	CAACGGTACA	ACTTTTTTCAG	5700
CAGATTCAAT	ATTTTTTTCA	AATAAATTGA	TAAACTCGTC	TTTCACATCA	AGCTTTCCTA	5760
TGATGGCAAG	TAATGTAAAT	AAGTCATCAC	CTTTTAGTTC	ACGCATTTCT	AAAGTCATTT	5820
TSTTATTCTT	CCATGTAAAA	AAGAACGACT	GATAAGCCGT	TCTTTAATTA	TTTCAGTTCTG	5880
TCTAAATTTG	TTATTCTGAA	ATATTGATT	CCTCAATAAT	CTACAAGCTC	TCAGTTAGCT	5940
CTATAGGATT	TTTCGATGTA	CTTGGGTAAT	AAATACGGAA	TGGGAATTCG	TCATTTATTA	6000
GCTGTTCAAT	AGAAGCGTTA	GCTGTGATTT	CCTGTTCAAT	TACTACCTCT	TGCCCATCCT	6060
CTGTTTTTAA	TTCAAGTCCG	CTTTTCACTA	ATCCGTTATC	TAACGCAACG	ATGATAGGCT	6120
GTTTCGTTCC	ATTATGAATA	CCCACAACGG	CCATGTTAGG	GATATAGTCC	CCTTCTTCTA	6180
AATAGCGTTT	AGGTTTAATA	ATTTTATAAC	CAGCAGGTGC	TTCATCTGTA	GTAGCATCTG	6240
TCATCGTAGC	ATTTAATGAT	CGGCGTAAAT	TTTCTGCTGT	TAATTCAATC	AAGTTGGCTT	6300
TCATCGTAGC	TGTGGCTGAT	TCAAGCACAT	TCAAACCTAC	TACGTCCATA	ATATAAGCAC	6360
CGTCTACTTC	CACTTTACGG	TAAGAAAGTT	CTGTTTTGAT	TTCTACACCA	CCAGAAGTTG	6420
CCCCCATGGG	AATACCGGTA	AACTCCTCTT	TCTCTTTATC	GTATTTAAAA	TCAGTAAATA	6480
CTGTCGCAGA	ATCAATGACA	AAATTGTCTG	CGGTTGTTTT	TGTATAACCT	GTTTTAGGTA	6540
GTGCCATATT	CTTTTTTCCT	CCAATCGACC	GCTACATAAA	AGCGGACATT	TCTTCTTTTT	6600
AACAGCTCTT	CTTTTGTGG	TATTTTATTT	GATCCTTGAA	AACTAAAAAT	GAGATTTAGT	6660
TCTGGAGTTA	GTACCCTTTT	GTAACAAAGA	GCTGTTTTAA	ACTTGTCTTC	TAGAACAAGC	6720
AAGTTTAAAA	ATGAATTATT	TTTGTCAAAA	ATATCAATAT	CCAGATAAAA	ACCATCTTGA	6780
TTTCTACGAA	TAGGCTCGGA	ATCAAAATCA	AAGGTGGCAT	ATGGATAGAT	GACCTCTTTT	6840
CGAGGATTCG	TTTCTAAAAA	TGTTTCTGGG	TGTATCTCTT	GTAAGATTCT	TACCAGCTCT	6900
CGTATAAACT	CTTCCATCAC	TTATCCCCCT	TAAATGTAG	CTCCGTATTC	TTTCCCTAGA	6960
ATTTTCGATGA	CTTTTTCTTT	ATTCTGCCGA	AACGCGTTTC	TTAAAACTT	TTGTGGTTTC	7020
TGACCTCTTG	TAAAATGCCA	TTCTCCATCT	GGTCCCTCAA	AGAACCATCC	ACCTTTACGT	7080
CCTGAGCCAT	TTTCAGCAAA	TTCTCCGGTT	CCGAATTCAA	CATAAATGGC	GTATTGTTCT	7140
GGCGAGCCGA	CAATGCCGAC	CACTTTTCCT	CCAACAGCCT	TAAGTTGATA	GTCTATCCGG	7200
TCCCGCAACT	GTCTGTTC	AACTGGCGCA	CCAGACTTTG	CAGCAGCTTT	GATAATCATG	7260
AGGACTTTAG	TCATCCCCCG	TTCTGCCGTC	GCGTCTAACT	CTTTTTTTAC	CTTGTCTTTG	7320
TAGCTGATAA	ATTTAAAGTC	ATTCTTCGCC	ACTTAAAAACA	CCTTCAAACT	TCAAATAAAT	7380
TTCATTGTGA	TGCTTTATGC	CAACAGGATT	ATCGGAATAC	GTAATATCGT	AATAACGATT	7440
TGTTTGATCA	ACCACTCGCA	TATTGTCTGT	TATCCCCTCA	ATAAATTCGG	GAATAATAAG	7500



AATGTGTGTA GATTCTTCTA CAAATGCGTT CTGTTTCGTT GTTTCATCTG TTCCTGTCAC	7560
CAAGTCGATA TATCCAGAAA CTTCTTTAAA CAAGCCCCAA TCGTCAATAA GCCCGCCTAA	7620
TCCGTCAGGC TTTTGGCCTT TGACCTCTTG TAGAAAAAAG GTTTGAGGTG AGTACATTAT	7680
CCCCACCTCA TTTTTTTATA TTTATTTAAA AAAGTGAAGT TTGCTGCAGG AAAGCCCTCG	7740
ATATTATnTCA CTAGCGTTCA CATCGTAATA CGTGATAGAC ATTCTAGCGA TAGATTCTGA	7800
CTTAATGCCT AACTTTGCGC CCATCGATTT TTTAAATCGT AAAAGTTCTT CAATACCCAT	7860
TTTAATATCT GCTGGATATT CTACTTTTGT AATAAAAAGCA CCAGAAAAAG AGCCATCAAA	7920
AAAAGGCTCT TCAACATCTG TTTCCAGTGT CTTTTCTGTG ATAGCTTCAA CAGTTGTTAG	7980
GCAATCGTTT ACTTTAGAAT TACTAATTTG AATCGTATCA CCTACTCGTA AACCAAGAGG	8040
GTTATCTTTT AACACAATAG AATTACCCTT AAAACAACA TCCCTAAAGC GTATATGAAC	8100
GTGTTGAAAA TTATTATTGG TTAGTTGCCT AACTGTCTGT TCGAACGCGT CTAAGTCACT	8160
CTGAGTAACC GATGGATCAA TTTTTTTTGC TTCTTCAATG CTGATAATCA AGTTAGCCAC	8220
CTCCTAAGAT GCAGGTGTAC CTGTTACTGC AATAGAGGTT GTAAACGCAC CGGAAGTAAA	8280
TTCAAATGTC GCTTTACCTT CTGCAATAAT TGAGCCATCA AATCCACCGT TTTTATTTTT	8340
AGTGATAGTA GCAACACTCT CATTACTAGA TATCACTGTA ATTGCTGAAA TTACAGCTTG	8400
TGCGTTTGTG GCATCTACTG GATTAGCTGC AATAGTAAAT GTCTTAGTTG CGCCTACTGC	8460
ACCCGTCCAT GTTTTCTGAC TAGGCGTGAT ACCCGTTGCT GGATTAGTAG GCGTTACGCT	8520
TTTTTTACTG CTTTTAAAT AGCTACTTTA TTTTTTTCAG GGATGTGTTT ACCATATTTT	8580
CCAGCTCCCT GGATTGCTAC TCCGTAAAA TCTTCGGAGT CAATAGCACG TGTTACAGAA	8640
ATACCAACGC CCGCAATACC TACGCTATCC GCGGAAAATT GAGCCATTTC CCCAGTTTGT	8700
AATTTGCTTT CTGGAATTTT AACCAATACA AAACCTTTAA ATTTATATAT TGTTTGTTCA	8760
TCAATGTTAG CACTTGAATT TTTAGTTGTT GTTGCTAAAC CATTGTCTAC TAAAAATCA	8820
TAAACATCAG GGTGAACATA AGCAACCCAA GCAAGAGACT TAGAACTAA ATTATTCACA	8880
AAAGTCTTAT GCGCCGTTGA GAACAGCTTA GTTACTCCTT CACTGGTTAA TTCAAACCGA	8940
ATCGTTTCAG AGGCTGCATC AGATAATGCT TTACCTAATA AACCGTCGAT ATACTCAGCC	9000
CAAGCAATCG CTTGTTCCCTC TAATCTTTCA GCTACTACTG CGTCAGCGTC aTCGTTTACT	9060
GTGATATTAT CCACTCCTTC GTGAATCGCT AACGGAGATT CATAAGGAAC CGTTGTATCA	9120
ATAGATTTAA TTTCTTTACG TGCCCCAAAA CGATTAGAGT TGCCAGTTCC TGTACCAAAC	9180
GCTACGTTTT CACCAAGTGT GTAGGGTTGC ATTACCACAG GTGTATCTGA AGTTTTTAGC	9240
ATCAAAAATG TGTCGCTATC TTTTACAGCA TCCGACGTCT GCAACTCTCC TCCAAACGCA	9300
CGTAAAAAAT AAGATTTTTT CTCTGTAATT TTGGCTAATA AGCCTGCATA TTGTTTCGTG	9360
TAAAATTTTG TAGCCATAAA AAATTCCTCC AATTACTCGT ATTTTGATAA GATTGCTTCA	9420
AATGGATCTT TTTCGTTTGG TTCCTTCCCT TTTGGCAGTT TGTTATCAAT TGGTTGATAG	9480

CCtGCTTTAT	TTTCAGTCTT	TTTTTTCGAC	TTATCATCGA	GTGTATCCTT	CTTTTCAAAG	9540
TAATCAGGGA	TGGAAGCTTT	CAAGTCTTTC	ACTTTACTTT	CTAAGTCCTT	CACATTCCCA	9600
TCTTTGTCAA	GTTCTAACTC	ACCTAATTTA	AATAAAGCAT	AATCAATATC	CTTAGCCCCT	9660
GCACTTTGTA	ATGCAGCAGA	TACTTGGCTA	TTGATTTTCA	AATCTTTGTT	TTCCTTTTCC	9720
AACGCGTCTG	CCTTTTCTTT	TAAATCAGCT	AGTTCTTTTT	GAACATCCGG	ATTGTCCTTT	9780
GTTTTAGCTT	CTAGCGATTT	TAAGGTTTTA	TTTGCTTCAG	CTAACTGGCT	TGAAAGATTA	9840
TTATATTGGT	CTTTTGGTAC	TGCATTTACA	GGAAATTCTT	TGTCAATTTT	TTTATTAGCA	9900
GCCTCCATAT	CCAACGTACC	ATCTTCTTTC	ATATGTTTTG	CTAAAATTTG	TTTGATCCAT	9960
TCCATCGTTT	TCTACCTCCA	TAGCATTTAT	ATAGCGGTCG	CTGCCGCTTA	GAGTGTCAGA	10020
ATATACCGTC	TGCTCGGTAG	CGGTGECCTT	TTAACGTCAT	GGCTCCTGGA	CAAAATAAAA	10080
AGACTTATCA	ATAAGCAAGT	CTTAACCTCT	CATATCATCT	AATACTTGTT	GTTTTGCTAT	10140
ATCTAGCATG	CCTAGCAGCT	GCAAACCTCC	ATTTTGACTG	TAATAAATAG	TGACCTCTCC	10200
ATTTGGCCGC	TGTACAACAA	TTGTGGCCTG	TTCAAATCA	GGCTCTTGGC	TTGATACCCA	10260
TTCGTAAAC	TCTTCGTAGT	CTACATCTAA	ACAAGCGTAA	GGATCTTCCA	TCTTATGGCC	10320
TCCCTTAGAT	ATAAAAATAG	CACTCAGCCG	CTTAATGGGT	GAGTGCTACT	ATTGCTTATA	10380
ATTATCTGCA	TCCGCTTGCC	ATCTATCTAG	CAAAGTGGGT	TGATTAAGTT	TATGTTGTTG	10440
ATAATATTCC	TGTCTCAACC	GCTCACGTTT	CTCAGGCGTT	TTGCTGTGT	CTAGCAGCTT	10500
ATATAACTCA	TCAAATTCCT	TATCAGTCAT	CTCTAAAGCG	CTACGCCTAT	AATCCACATC	10560
TACCTTACCC	AAGAATTGTC	ACCTCCAATA	TATTATCGAC	TATCTCCACT	GTATACTTCA	10620
AGTTTCTGTT	CAGTAGGAAT	TCACTTTCCT	TTTTCATCGA	ACTGTTGTT	CCTATATAAG	10680
CGCCTTGCGT	TCCTTTAGGA	ATCAGGAATT	TAACACTCT	ACCATCCTTA	TTTGCACTGG	10740
TAAATTTTAA	AAAATCCTCG	GCAACTTTTT	TGTTGATACT	AGTAGACTTG	AATTCTTTGA	10800
ATACATTTCC	GTTCAGAATA	GCATCGTATT	CATTTGCAGA	TACACCTCTA	TAAGTTATTA	10860
TATCATGTTT	CAATTTAAAT	TTACTAATAG	CATCATCTAA	ATTTTGAATA	ATATTGACAA	10920
GATAAGGTTT	TTCTCTTGGA	TTATATTTTT	CATTTGCGAG	TATCTTATTC	ATTTTAGCAA	10980
ATGCGTCTGA	TGTATAATCT	TTCATAGCAT	TAACCTGTTT	AGGCTTCAAT	CCAGCAACCC	11040
ATTTATCACT	ATCATCTTGC	AGATCAGATT	GACCGAAGAA	GTCCAGCTCA	GTAAAATTCG	11100
GCTTTGGTTT	CGATTCATTT	TGTCTGATTC	TATCTGCATA	CCACTCATCA	TAGCTTTTAA	11160
ACTCCGACAT	CTCTTTAGAT	TCATTATCTT	TTCTAAGTTC	AGGGCTGATA	CCGTCGACAA	11220
CTTCAATTGT	TGTACATCGG	CAATTCACAT	CTTCTGATGC	TACCCCAAAC	ATCCGCGGAC	11280
CTTTTGCCGA	ATGCCCTCTA	ATGGTAAACT	CTTCGTCTAC	ATCAACCTCT	TTACCGTCTA	11340
ACTCTTGGTG	GGTTCGGCGT	GTATGTTTAT	CGATCGTGGC	AAGCCATTTT	TTTTTAATAT	11400
TGATGCCCAG	CTCTTTTGCT	TCTTCATATC	CTTTTGGGT	AGTGACAGAC	TGAGTACGTC	11460

CTGCTTCTGT	TCTTGCAATA	CGTAATGCTT	GTTTGTAGCT	AGCTTCTGTT	TCCTCATTAA	11520
TCCATCTAGC	TATTTTCAGCA	TAACTTTTAC	CCTCGAATAA	GCCCGTTATG	ATATTGTTAG	11580
TCACATTTTG	GGCTAATTCA	TCACGGTACT	TGTATAAACG	CTTTGAGAGC	CTCTTACCCG	11640
CTACAGGTGC	ATTGACGAGA	TTCATAATAT	AATCATGGTT	AATTAACGGC	ATACTTAGTG	11700
CTATATTCTG	TGACTGCTCT	AGCGTATACC	ACAGTCCATA	ATATCCCTGT	TCAGCTTGTT	11760
TTGCAGAATA	GCCTTTGATA	GTTTTTTCAA	CTTTCGGTGA	GTTTAACTGA	AGAATGGCAT	11820
TAATTTTCATC	AGCAACACTA	AAAAGTCTCT	CTACTTCCAA	ACGGGTGGAA	AAAGAAAGAC	11880
TTTCTGCATT	TTCTGTATAA	ACTTTTAGCC	GCTTTTTGAT	GTCGATTAAT	GATTGACGAT	11940
AAACATTAAA	TAGCTGATTA	TCCGTTTCCT	GGTAATTGGC	TTTCTGTAGA	GCCTGTAGCT	12000
CCTTCTGCCA	TTTGTTGAGT	TGGGACATTT	GGGGTCACCG	CCGTTTCTTC	TTCATCGGTG	12060
TTGTCCGTCA	AACCTTGCGT	ATATTCCGCC	TCTTCTAATG	CCTCTTTCAC	GTCTTCCCAG	12120
TCAAGGTCAA	ACTGTTGCA	GATAAGGCGT	AAAACGTTGT	CGTCGTCAAG	GCGAGGTGCT	12180
ACTTGCAAGA	TAGATTCAAG	GATAATTTTT	CTTGCTCTG	CTTCGGTTTT	CTCATTGTTA	12240
ACAATATCTG	TTTCATTAC	CATTACTTCT	CTAGTAAACG	TAAATGAAAC	TTCTGTTGGG	12300
TCAAATGCTT	TAGTATAACG	ACGATTAATG	TCATCAATAA	CAAGCTTATT	CATCCATTCT	12360
AGTAAAGCTC	GTAGTCTGGC	TTCAGTCTTG	TTGGCTTTCA	TGTTTAATAG	TGTATAACGT	12420
GCTTTTATCA	CTATGTTGGT	GATATTCCCG	TCGCCTACTT	GAGTTGAATC	AAAGGCCATT	12480
CCAAACTTAT	AGATATTTTC	TTTGTCATC	TCCATTTTGG	TTTTACGCCC	TTCAGTTGGA	12540
ATCGTTACTG	TTTTAACATC	TAAGCCCCCA	TCTGAACCAG	TACCAACAAC	CTTTTATAGAT	12600
TTAACATTTT	GTCTTAATTT	TGAAAGATCG	TCACCTTGGA	AACCTGATAC	AACGTAAATG	12660
GCTTCCGCAA	AATCTTGTA	ATTGTTTGAT	AAAAACAAT	TCATTAAGTC	GTAATCATCT	12720
ATCAGTGCCT	TAATAGACTT	TAAATCCGTC	GTTTCCTGTT	TATTATTTGA	TAACCGATAA	12780
AAAGGTATTT	GTCCGTAGCT	TCGTTGCAAT	AAGCTTTCAT	TATCGCTATC	AACTGCTAAA	12840
ACATGTGGCC	TTGGATTAAT	TGGTTCAGCT	TCATCTAATT	CATAATCTTT	GTTATCCTCA	12900
GCTACGAAAA	AATAAACGTT	CTGGTCAGTC	CACACTTCCG	CGTGATGGAT	ATCGACTGTC	12960
TCGCCGTCTT	TCTCGATTTC	AGTGATATAA	TGACGGCAAA	TACGTTGTAA	TTCATTGTAT	13020
TCATTGTATA	CACCAAACAC	GTTTAAACTA	TCAGCCACTT	GAAAACACAA	TCGATCTTCT	13080
GCATTGGTTC	TTGCATAAAC	ATACTCAAAA	CCTTTTTGGC	TTGAACCTTC	CACAAGCTCT	13140
TGTAGCACTA	CTTGAAATTC	AGAATTGTAA	TACTCTGCTA	AATACTCTTT	AAGTTCTTCG	13200
TTTTCTGTTT	CGTATTCAAC	AGGATTAGAT	AAAAGGTATT	GAGTTTTTTG	GTCAACAATT	13260
TCTGGGAAAA	AGCCGTGCGG	GATTCGTACA	TTCGAGGCGT	ATTTGTCTTC	TCGCAGTACC	13320
CCCTCATCGT	CTACGTAAAA	GATACGATTG	TTCATGATAT	CATTTTCATG	GTTATAGTAC	13380
CGAATACCTG	TCTCTGCCTC	TCTTTTTGAG	GTAGATTTTC	GATCTTTGTC	AATCGCCGCC	13440

TTTAAAGCAC TAGCGATAAT CTTCACGTCT TCACTAAGTA AAGCTTCCAT TTACAGCCTC	13500
CTTTCTAGTA TAGCCATTTA TTTGTGTTCT TAATTTCTCT AAGCAGACTA GCTGCACTGT	13560
CTGGTGCGTC GTCGTGCTCT GCGTTCTCTG TATAGTCTAG TATCTCGGCT ATATATTCTT	13620
TGTCTGTGTC TTCAAGCCAA ATTATTCCGC TCCAATACTT CCGCAAATAA GAAGATATCT	13680
TGATAAACTT ATTTGTTTTT TCGTGATACT TTTGCACGTA TTGACTTCGT TCGATTAGGT	13740
GTTTAGCTAA ATAACCTTTA TCGCCGTTCC TTTCAGTGTA AAATGTTCCA GCTTGGTAGT	13800
ATTGATGTAG TTGCAATATT TCAGGTAGGC AATCATCAAC GTGTTTCTGC CATTTCTTGC	13860
CAAACCCGAT AATGGTTCGG TCCTTTTGCT CTTTAAAAAT GGTAAATGCT GTGCTATCAC	13920
CACCTCCGTA TGCCGCGTCG ATGTGTGCTA CACCGTTATA AATAAGGTGA GTGTTATCGA	13980
TATAGGTCGG TGCAGTAAAT AGCGATTCAC TATCAGCGAT GTGTTTTAAC TCGTAGTTCG	14040
CTGCAAAGAG CGACGGTGTC ATAGACTGCT GTAACGCTTT ACGTTGTTCT TTGTCAATTA	14100
ATCCTGTTTC GTAACAATCG AATTTCTTGA CGTTAGGCAT TTTAGAAATT GCATCTTCCT	14160
TATGCCAAGG CGTGCCTGTG TTAATAAAAC GGCCCTCACG ATTCTTCACA TTCTGCAATT	14220
CTTGATACTG CAGCTTTGTT TTCTCTCGTT CAGCAGACT TACTCGGTCT TTAATGTTAA	14280
CAATATCATC AGTGATAACG ATATCCGCGT GTTTACCTGT TAATGAAGCA TAAATCCCCA	14340
TACCAAGTAA TTGAGATGTT CCCC GG GTAG ATGTTTTTAA GTTGGTATCT ATTTCTGTTG	14400
TAGTCTCTTT CAAAAGCACC AATTCAACAC CATATAATGC AAATACAAGC GTTTTAAAGT	14460
ATTTGCTAGA TAAACCTTA GCCACTTGTA AAATAATCTC TACGACGTCT GTATCGGTCT	14520
TACGTAAGAA GATAATATTT TTGTTGGGAA AAAGAACCAT CAACAATGCA ATAGCAATCG	14580
CCAATGTGGT TGTTTTAAAT GAGCCACGAT GAGCCAATAG CGTTTGATCG TCTTTCTCGA	14640
ACAAAAAGA CTTTAGCCAA TCATTGTGCA GTTCTGCTAA GTCGGTAAAC CCAACTAAAT	14700
TTCCGAAAAT TACCGGGTTC GTTTTGATTA AGTTCAAGTA TTTTCGCTTT TTAGGGCTCA	14760
TGTAGAATCA TCCTCGAAAA ATTGCTCGAT TTCGACAGCT GCATCGGAGA TATTAAGCAC	14820
GCCTGAAACT TCAGTTTCTT TTCGATCTCG CCATTCGTCC GGCTTTCTAT TTTCAACCA	14880
GAAAATAGCT GCAGTTGGAT TAGGAGCTAC TTGTTTTTTA ACCTTTTTTG TAATTTCCAT	14940
ACCGTTCTCT GTTAGTTCTT CTGTAAC TTCGTTATTCG TAGCCTGTAG CACTCTTAA	15000
TAAAGCATTT TCAACTTGTC TATCAACGAC TTCTTTTCCC TTTTAAAGG ATGCCGAAAG	15060
TGCCGGAAAT TTTTACCC ATTCTCTGAA AGTTGAATAA GCTACCCCGA TATTCTGGGC	15120
TATCTGCTTA TCGATGAGGC CATCTCGTGC CCATCCTTCG ATTTTGATTA ACCCTTCCTC	15180
GGTTAGCCAC TCTGTGTA CT TCGCCATGAC CTCACCGCCT TTCTGTATCA AAAATAGACA	15240
CCTTCCCAT TTTCTGGGAA GGTGCCTTTT CTGCGTAATT TTAGTGATAA CAGTATATCA	15300
CGATTTATCA ATGGGAATTG TCACATCATT GTCACTTTTT TCGATTTTCT TCAAAATCTG	15360
GTTATGTTTA TTTCTGATAT AGCCATAGCT ATGCCCAAAT TCTCCCGCAA TTTGTTTTAA	15420

CGATAGCCCT	TCGATATACT	TTGCTACTAA	GATTTTTTTGA	TCTAATCCTT	TGAAACTATA	15480
AACTAGCTTT	CGTAAATCGT	AAATTGaGTT	CATTTTCCAC	GCTAAACGTT	TTTCTAAGTC	15540
TTCGATAATA	TCTTCCAAC	TAGAAGCTTT	GGAATCTTTT	GTTAGTGTGT	ATCTTCCTAA	15600
ATCCTCTGGG	TCACACCATC	TTTCTAACTC	TTTTTTATAC	GTTTCTAGCT	CCCAGTCTAA	15660
GTAATAGATT	TCTTGTTCTA	ATTTTTGATA	GCTGACTAAC	CATTTCGTACA	ATGTACTTTC	15720
ACCTACCTTC	TTGTTTCATCG	CTTTCACTAT	TAACCGTAAC	TTGTACATTT	TTATACAGAA	15780
TTAGAGTTAA	TAATGCGGAT	AGCAATATAA	ATCTAACATC	ATACGGTATC	TTAGCTAATA	15840
TCGAAATAAT	CCACTCCGAC	ACTTTGTATA	TCATCATAGC	TAAACCTAAA	CATAACCAGA	15900
ArGATATTAA	TTTATCCAAG	TTTATCCCCC	CTATACTCCT	CAATTTTGGC	TTTCACTGCA	15960
GCCATTAGCG	CTGATTGCC	TTGTTCTTTT	GCTTGTAAG	CTTGATTAC	TTGTCGTCA	16020
ATAGTTCCTT	TCGTCATAA	GTGATGAATA	ATAACAGGCT	GTGTTTGTCC	TTGGCGGTCT	16080
AATCTGGCAT	TGGCTTGTTG	ATAAACTCC	AAAGACCATG	TAAGCCCAAA	CCACACGATA	16140
ATGTGCCCGC	CTTTTTGTAA	ATTTAGTCCG	TGCCCTGCTG	ATTGTGGATG	GGCTAAAAGT	16200
AAAGAAATTT	TTCCCTCGTT	CCATTTCTCA	ATATCCCCGT	CAGACACATT	TAAAGCCTTT	16260
GCTTGTTTAA	ATCGTGCTTG	GATTCTTTCT	AGGTCATGTT	GGTATTGATA	AAAGACTAAA	16320
ACCGATTGCC	CTTGTCGTC	CTCTATAACG	CGTTCTAGCG	CATTAACTT	TTCTTGGTGT	16380
ATTTCCCTTC	CGTCGCCGTT	TTCGTCGTAT	ACAGCGCCGT	TAGACAGCTG	TAAAAGCTTG	16440
TTTGATAGTG	TAGCCGCATT	ACTAGCTACA	ACGTCTGTTT	CTTCGAGTTC	TAACACGTAT	16500
TCCCGTTCTA	GCTCTTTGTA	CTGTTTCCAG	CTTGTCGGGT	TCAAGTCTAG	CTCGATAATA	16560
TTTTCCGTTC	TCGGTGGCAG	TCGCAAATAA	TCTTTTGCTT	TCATACTCAC	GCATATATCG	16620
CTTATTTTGT	TGTAAATCGC	TTCTTCTGCT	CCTGGGATTA	ATTGCCAAGA	ATACACGATA	16680
TGCCCCGTTT	TTTGTGCGGG	GACAAAATAT	TTATTACGGT	ATTGTGTGAT	AGTCTTTCCT	16740
AGCCGTTCGC	CTTGGTCCAA	TAAATACATC	TGCGCCACAC	AATCCATCAA	GCTGTTTGGG	16800
GAAGGTGTGC	CTGTTAGCCC	GATAACACGT	TCCATCTTCG	GCCGTACTTT	TCGTAATGCT	16860
TTAAATCGTT	TGGCACTACT	TGACTTAAAA	CTGGATAACT	CGTCTATAAT	CACGGTTTTA	16920
AAGGGCCAGT	TTGTTGATA	GTAATTGACT	AACCACTCCA	CATTTTTCACG	ATTGATTAAA	16980
TAGACATCTG	CCTTTTTGAA	TAAAGCTTCT	TCACGTTGCT	TTGGGTTTCC	TAACACTTTT	17040
GAAAAAGTGA	GGTGTGCAG	ATGGTCCCAT	TTTTCAATTT	CATCGGTCCA	AGTTTTCTCG	17100
GCAACTGATA	ATGGGGCAAT	TACTAACACA	TTTTCAATAA	TTTCAAACGT	GTGTAACAAC	17160
TCGTCAATAG	CTGTCAAAC	GGATAAGGTT	TTCCCTAGTC	CCATATCTAA	CATCAAAGCA	17220
CAGTAAGGAT	GATCAAGAAT	AAAATTCTTA	GAGTATTCCT	GATAGGGATG	TAACGTTGCT	17280
TTCAAACCTCA	ATCACTCCTT	TTTATGGGAC	ATATCAGGCA	TAACCTTGC	TATACCGCTT	17340
TGCATTGCGT	AGATTAACAT	ATCCACTCCT	TGCTTGCTGT	CAATGACAAA	CACTGGGATA	17400

AGCTGCTCTT TGaATTCTC TATAATTTT AACTGGTCTT TTCTTGGTTT ACCTCCAGGC	17460
CTCTTTAATT CAACGAAAAA CACGTTGCCT TGATATAAAA TAATCCTGTC aGGAACACCC	17520
CTAGTTCCTG GGGAAAGTGA CTTATAACAT AAAgCTCctA TGC GTTTTAT CTGCCTAACT	17580
AGATACTTTT CAATGTCGTT CTCAATTGTC ATTACTTCGG CCTCCCATCT GATAAGATAT	17640
TGGTTGATGT GTTCCATCAG ATCTATTTTC TTCAACCTCC GTTTAAAAAT ACCGTGTCAC	17700
ATGTAACACA AAAATTCTTC AACTTATATA TATCCCCAT GTACATATAT ACATATACGC	17760
CCTATATATG TATATTAATT ATTTCTATAT ATGAATAGAA AATAGTGTTA CAGGTGTTAC	17820
GATTGTnCTT CATTCTTACT CTCCCAAGAT GTAATAGCGT AACACTCCCT GTAACACCAA	17880
CGGGAAATAA GGTGTTACAG GTGTTACAGT GAATATTTTA GGTGTAACAC TCGTAGCTTT	17940
TTTAGGTGTT ACACCTAACC CCGCGTCATT ACTACCTTTT TGTTTTAAGG TGTAACACGT	18000
AAATAGGCTA CTTGTACTCC ATACCCAGGG CCAAATCTTA AACGTCCTTT ATTACCTTTG	18060
CTGTTTTTCT CCCAACCAGA TAAATGGCTT AAAACTTGAC GAATTTCTGC AGCTTTTGCA	18120
GGATGTATGT TTTTAGAATC TCCGTTATAC AATTCATTCC AAACCTCTGC AATACATACT	18180
TTGTTTCTAA CAATCTCGCC TTCTTCCTGA ATATCCGTCC CCCACCCCTG AATATACTCT	18240
CGTCTTTCTT GCTTACTTCG TTGTACCAA TCTTCCGTTA TTGGGATTTT CAAATATTCA	18300
AGGATTTCTGC CTTCCATGCT AGAGGTTTCC GTGTGCATTT CTTGGGCTTC TAGTGCTAGT	18360
TTTTCTTGTT CGTCTGTTAG ATACAAAGGT TCTCCTGCTT GCCACAACCTC AACGGCTTCT	18420
GcCCATATCT GATTGCGCGT TTCGTGCGTC ATTTCCACACA CTTTATTTTT AATCGGTTGA	18480
ATGCCGACAT CTA CTGGCCA AAAACGCCGA TTTCCTGTTT TATCTCTTAG AAACCTCGTTG	18540
TCGTTCGTTG TTCCCCAAAA GACACAACGG CGTTTAAAT AGGACTTATG CCGACCGTAA	18600
GCAACGCGGA AAATGTCTTC TTGCTTGCTA ATAAAATGCT TG GTGGCTTC AATGTCTGCT	18660
TTTTTTGTCG CTGACAATTC GCCCATTTCC ATGATCCAAA CACCTTGCAA CGCTTCGTAA	18720
GAGTCTTTCC CCGTCACACC TTCTAAGCTG TTAGAAAACC AGTCTCCCGC CAATTTTGCA	18780
GGCAATAGCG TTTTTCCTAT TCCTTGCGGA CCGCTAGTCA CTAGCATGTA GTCAAACCTA	18840
ATCCCTGGCA CAAAAATACG CCCTACGGCA GCTGTAAGAA ATTTTTTCGT AACTACTCGA	18900
ATGAACGGTG TATCTTCTGC GCCTAAGTAG TCTATTAGTA AAGTTTCTAC GCGAGGCACT	18960
CCGTCCCAGT GCAAGCTCTC TAGGTATTCT TTTACTGGAT CATATGAGTT TCTTTCGATT	19020
TCTTGACTA GCGCATCATC GATTTTGCCA CGGTTCACAA TACCGTAAAT TTTTTCGATA	19080
TATACCCGTA ACCCCGCATC GTCGCTGTCC TTCCACATCT TGTCTCGATC TAACTTCCGC	19140
CAAGGCAAGT TGTCTTTCAC CTCAATACGA TTAGAAAAAC TGTTCATAAA AATTTTCTTT	19200
TTGAGGTTTCG GGTCATTCAG CATAATGACT TCTAAATTTT TCGCGGAAGC CATAATCTGA	19260
CCGTATTTCG CAATTTTCGAG GTCCAATTTT GTAAACCAGT CTTTGTCTGC CTCTTCCAAT	19320
TCGTCTAGCT CGCCGTCAAA GTCTTCTAGT GCTTGTGACA GCTTTTCACT CTGAATCAAG	19380

GTCTTAACTT GCTTGTCTTC CATTGCGAAC TCTCGCATCG CTTTAAACGA TGGATAACGG	19440
TCAATCCGCG TCGTAGGTTT GACGTCTTCA TCTAAATCGC CAAATAGATG GATACGGACA	19500
AGGTCAAACG CGTTCACCAG TTGATCGCCT ACAGGGTCCG TACCGTGATG AGAATAAGCG	19560
AACTTATCAT CATAAATCAC TAGCCCACCA CTAGTAGAAC CTTCTAAAAA CGTCCAACGG	19620
TCCTCTCTTC CTGTGGACCC GTAGATGTCA GGCAAAAAGG TTTCAATCGC TGAAATGATG	19680
TCATACGTGC GACAAAATGC CCCGACAATC CCTTTTTTCT CTAATGGGTC GCCTGCTTTT	19740
TTTGCTTGTC TTTGCGGAT AGAATGCCCA CGGGAACTTT CTGGCCAAAA ACTTGCATCT	19800
CGCCAGTCTT CGTATTGGCT TAGAATTTCC TCAGGATCGA CCCACGGCAA ATCGATGTTA	19860
TCGGTGAAAT ATTCCCCGTC TATCGAATGA CTGGGCCAAT ACATTAAACG TTCTGCCTGA	19920
TAGGTCGTAT CATCGAAGTT ATCCATGCCG AAAAATCCG CAAGCTTTCT CGCTAGGGGC	19980
TCATACTCTT CGGCGGTAC TGGGCGACTT AACGGAATAA TCAAGCGATA ACGTGGACCT	20040
TTTACTAAAT GGCTGTGTGT CGTGTAGACA GCTGCAGCGT GGTCAAATAA CAGCTGTACA	20100
TCGTCCCAA AGTCTAACGT CGTGCTGTCG GCATCTAACG TAACAAGGcT ACGCTGTTGC	20160
GTGTTTCCTC TTTTCCGTCG TCCTTCTTTT AACCAACCAC CAATAAACGC CCCAACGTCT	20220
TTAACTTCGC CCTGCTTAGA CTTTGGCATT TTTTGTAGT CTTGACAGT CTCTTGTGTC	20280
ACTGTCGGCG TTTTCAATCG TTGGGTAAAA TCCGACCAAG ACATCTGCCG ATTTTCCAT	20340
TTTTTCTCGG TTTTGAAGC ACCCACTGCA AGATGGATT T CACCATCATA GGCTAATTG	20400
ATGTTTTTTT CTGGCTGTTT CATCTTTCCG CTTCTTTCT TTTATTGTTT AAGGATGAAA	20460
CCTAAACAGT CAAGCCATCC TTCTTGATAG TCACACGTAG AGGATTTATG ATCATAATTA	20520
AACGCAGCTC TCATTGCTTT AAGATATTTA CGCTTACCAA CAACGTAGCC CAACAAAAAG	20580
CAAAGTACTA AGTTTATTAG TATAAGTATG TTATTTCCCA TTTATTAACA ATCCTTTCCA	20640
TATTCGTA CTGTAATTAG ATTAGTTAAC CAGCTATAAT CTAGCTCTG CTCTTCGTCA	20700
AACTCCTGTT CGTTGACATA AGCAGCAAAA TCTACGATAT TTTGATATT TTCCTCATCA	20760
TAAAAACCCC AACTCCAAAT GAATGCTTTA ACACGAACGA GTTCCCCAAT ATCTTTTTTA	20820
AAAACAAAA CGGGCTCATT GACCATTTC AACTCATAA AAAATTACA TGTCCCCACT	20880
TCTTCTTCGT GAGGATTAGT TTCAACATCA ACTAGTTTGA GTTTCATTTT CTATCGCTCC	20940
TTTTTCTAAT TTTTAGTGAT TTTCAATTCC CAATCATCGC TATCAAAAAA TACAAATTCT	21000
AATTCCTCAA AATCGTTTGT ATTTTATATC AAATCTTTAA ATTTCTTATT TGAACAAATA	21060
TTTCCAAAGC GTTTTAAAAA GCGGTGCTTA AAATTTTCAT AATCACTGTA ACTAAATTCA	21120
AAAACTTTCC CAGTTGAAAA AACTCCACCA GTTGTGAGT CTAACCTTAC ATAATTTAGT	21180
GCCATTATCT GAACCTCCTA ATCTTTCATG TAATATTTTC TTTCAAATCC TTCGGCGTTT	21240
AACGGCAAGC CTTCGCCCCA ATCAGGAACT ACAGACATGA TTTTGTTTAT TTCTTCAATA	21300
GATTTTTCCC CTCTGGCAC TTCGGCTACT GCCTCATCGT GCACATGGAA AACAATAGGA	21360

TAGCCCTCAG	GTCAGCCT	CTGCATCGCT	TCTGCTAAAA	CGTCTCTTGC	CGTTGCTTGG	21420
ACAATGTTTT	CTACCAATTT	ACCGCCGTAG	GTTTGCTGTT	CCGTGAAAGC	GACTTTATCG	21480
CCTTGCCCTT	CATAAAAAAT	AGCAGGACCG	TAGTCACCTT	CTTTTAAATG	CGCTTTTGCA	21540
TACGCTAGTT	TTCCTCCGCT	TGGTAGCTGA	ATAAACAAGA	ATCCTGCTTT	TTTATAAAAT	21600
TTCAACCCAG	GAGGSCCTTT	TTTAATGCCG	CCGTTTTGTA	AACAGTCAAT	CACAGCTTTT	21660
TGGGTTTCAT	GCCAAAAGTT	CTTAATGCGT	TTATTGGCTG	TACGCCAACG	GTCCACAATG	21720
TCTTGTAATT	CGTGTTCTTC	AATGCCGTTT	TCTAATGCAC	CCATTGCTTT	AAGCGCCCCT	21780
GGTCCGCCCT	GGTAACCAAG	CGCCAATGTC	GCCACTTTTC	CACGTTGGCG	CATGTCTTTA	21840
CGTTCTGGGC	TTTTCAGTC	GTAGTCCGTC	ACTTCGCCTA	AATGGAACAT	CTGCGCCGCT	21900
GTTGCTTCCT	AAATTTTGCC	GTGTGTGCGG	AATACTTCTA	ATACCCAATC	TTGTTTGCCA	21960
TACCAAGCAA	TCACTCGGGC	CTCAATGGCT	GAAAAGTCAG	ACACGATGAA	CCGATGCCCT	22020
TCTTTGGCAA	CTAACCTGT	TCGGATAAGT	TGTTTCAATG	TGTCTGGCAC	ATCTTCATAC	22080
ATTAATTCGA	TGCCTTCAAC	ATCTTTTGCT	TTCACAAGCT	GACGGGCAAA	ATCAATTTCA	22140
CTTAATAGT	TTCTAGGCAA	GTTCTGTACT	TGTAATAATC	GCCCTGCCCA	CCGTCCTGTG	22200
CGGTTGGCAC	CGTAAATTG	TAAATGCCA	TGAATGCGAT	TATCTGAACA	ACGTGCATTA	22260
TCCATCATCA	GATATTTTTT	CGTGCTTGAA	TTAGACAAAC	TAAGGCGCAA	TTTAAATACC	22320
TCCGCCACAT	TCTCGGGCAA	GTTTCCTAAT	GCTAGAGCCT	TTAACACAAC	CTCTTTTCCT	22380
AACTTCTCGA	AAGGCGTTCC	TTGTTCTTCT	AGCCACTTTT	TCAGTTGCGC	TAAGCTGTTA	22440
GGATTTTCCA	ATCCTGTAA	AGCCTTCATT	TCGTTTAATC	CCGCTTCGCT	TAAATCGGCC	22500
ATAATATCAA	TAGCAGCCGT	TGCCAATTCT	TGGTCAATTT	CTGCCCCTCT	GTCGTTAATC	22560
CGCTGATCCA	AGGTGTAAAG	CTTCCACTCA	CTTTCGGGGA	CAGGGAAGCG	ATTCAACTTA	22620
TTTCGCGATg	CCATTTCTAC	ATTGACGTCT	TGAATACAAT	ACTCTATAAA	CGTTTGCCAT	22680
TTTTTCGGGAG	CGTGTTCAGG	TAAATTCCTT	GTACGCATGC	CGTTTTTCTT	GGTCGGCTTA	22740
CACGGTTTAG	AAAGAAGTT	AATCAGTTGC	GTACCTCTTG	TGTCTTTTGT	CTGTTTCGATG	22800
TTTAAATATT	TCGCACATTG	CCCTAAAGAA	GCTGGCAACC	CTAATTCGTT	TGCGTGAACC	22860
ATCGTACAGT	GCCATTGTGC	AGGGTCTAAG	TAGTACGGAA	TACCTAAATA	CTTAGACAAA	22920
CAAACGCGCT	CAAAATTGGGC	ATTAAATGCA	ATTTTAAGCA	CGTTCTTATC	TGTTAGCGCA	22980
GCCACTATCT	CGTCAGGTAT	TTCGCTAATT	GTAAAGTCCT	CACATTCGAC	AGGCTGACCA	23040
TCAACCGAAT	AAGCAAAAAG	TAAAATTTCA	AAGTTGGGGC	TATCAGCGTA	TTTATAAACC	23100
CCAACCTTGG	TTAAGTCTTC	ATCCGAATAC	GTTTCAATAT	CAATATTTAA	CGTTTTTCATT	23160
TGCTATCAGC	CGTTTCAGAT	GGTAACCGCA	ATTTACTTTT	ATCAACGGCA	ATCATCGGAA	23220
GCTGTGCTTC	TTCTCTTTA	AAAACATCAA	CGTTGATCCC	TAGCTCTTCA	AGATAAGCCA	23280
CTCCTTGTTC	AGCAGGCACT	CTTCTACAT	TTGCAATAAT	CGCATTGTAC	GTTTCTCTTA	23340



CCTCTTGGAT TTCTTGATCG TAAACAGCTT CTACTTTCGC AAGCCCTTTT ATTGAGTTAT	23400
ACTTAAGTTT TTCAAGAAAC CCTTCTTAT AGGTTTCAAA ATCTCTAGCA TCCGTATCGC	23460
GATAAAAGCT CCATATTTCC AAACGTGTTT CTTCTAACAA ATCGACAAGT TCATTTTTTA	23520
CGTTTTGGCG AATTCTAAGT GAGCTGAAAT ATCTATGCAA TTTCTCAAAT TCAGGATCAA	23580
TTTTTCAGCCA CTCTGATTCA ATATGTGCAC GAATAGCCTt CGTCTTTTTC GCTCTTATTT	23640
CTTTTTCCAC TGAATTTTTT AAAGTAATCA CATGGTTTCT TAAACGTGCT TTAGTTGGTT	23700
TCATTTTTTC CTACCCCTG ATCTTCTACT GGTGTATATT CTTGTTTGAC ATAAAAATTA	23760
GTATCTCCGA ATTCTTTGAA TTCCCAACGG TGCAATCTGA CGGATTCTAA ACGGCTACCG	23820
TTTTCAAAAT ATATAGAATC CAGTTTTCTC TCTTCAATTA GCTCAAAAAC ATCTTGTAAT	23880
TCAATAGGCA AATAAATTGG TCGTTTACCC ATTTCAATTT CTCCTTTCAA AAGAAAAGGG	23940
GCATTTTAGC CCCTTATTCA ATTTTtagCT AAACATATCG TCTTCTTCGT CTTCCTATTC	24000
TAGGTCACCA AAATCGGATT CAGCATTGGC ACGACCGCCT AAGAAGTCCC CTTTACATAA	24060
CGTTAAAAATG TTGTTTAGCC CAGCAGTAAC CCCTTTGTTT CCTGcAGTAC TGTAAGCATA	24120
AAAGTTAATA GATGCAATCG CATAAACACC TGAATAGACT TCGTCTGGAT CGTCTGTTTT	24180
TACAAGCACA CCATCTTCAC GTTTTACGAC TTGTGGTTTC GTTTTACTTG ATACGTTGAT	24240
AAACATTGCG TTTTCAAATT CTGGGCGTTC TTCGGTGTCC ATTTCTTCAT CGCCATCGCG	24300
TAGtAgTTTT CAAGCGCTCA AATTTAACGC CTTTAAATTT ATCCCCTTTT GCGCCTTCAT	24360
AGGCTGTTTT GATTGCTTCT TTCATCGCTT TAAGGGTTTC CTTATCGTCT TTTGGAATAA	24420
TCAACATGCA AGAATATTTT TTCTCTTGCC CTTCTTCCAT CGCATGAGGT TCTAACACAT	24480
GCACAAAAC TAATCTCACT TGATTGTAA TTACTTTCGT TCCAGTTACT TTTGCCATTT	24540
TAAATTCCTA CTTTCTTATA TTTTTAGTT AAATAATTCG TCTTCTTCTA AATCATCTAA	24600
CGAATTAGAA GCTTGTCTG TAAAATACTG TTTAATGTCC GAGATGATTT GTTCTTTCAT	24660
CTCTAAAATA GATTCTTTAT AGTTTTTTAG GTTCATTGTC CACGTTTCAT CTACAAGCAA	24720
TGAGTTAGTA ACTAACGACG ATAACGCCGC TTCAATCGTG TTAATAATATC CTTTAAATAA	24780
CACTTTATTC TCACTGAATA CATAAAGCCC AACATTGCGT GTATCTTCTC TTGTGATGTA	24840
ATAACCTTTA CCGCTAAAAA TCACAGGCTT TTCTTTACAT GATTTTTGAG GCATTTTATC	24900
CCCCACCTTC TACTTAAATA CATCCAAAGT ACCTTCTTGA CGTTCTAAGA TAACCACCGT	24960
ATCTGTCACC TTAATATTTA CAGGGCGATT TCCTGTGCTG TTTTCAACGG TAACAATAAC	25020
CGCATTTAAA TCATCAAAAG GCTGTATGAC AGTAAATCA TCCTCAGaTA CGTTAATTCC	25080
TGAATGCTCC GCTTCTCCA TTCCCTTTTT TAGAGTTGAA GCTATTGCCT CTTGTGTGTC	25140
AATTTTTTTC ATAGCCATTA CTCCACCCCT TCAAATCAT TAAGCGCATT TTCTACGCTG	25200
TTTAACGCTG GTCGCTTGTC ACTTkCAGGA ACAAGAACAG GCTTTCCTTC TGGTTTGATA	25260
ATGAAGTCTG CAGCTAATTC CGCAAATTTT TTCTTGCCGA CAACTTTTTTC TAATTGCCCA	25320

ATGGCTTTCA GTTCCTGCGG TTTTAAATG TCTTCATCTT CAAATCCTTC GGCTTCTAAA	25380
ACCATCAGTA GCCCTTCATT ATCTGAAATT TTTCGATTAC TtCGCCCAGC AACTACTTTC	25440
CAACCAGGGA ACTCTTCGCC TTCGTCCCTT GCTTTCGTGA AAGCGTAAGT TTCCACATGT	25500
TCCAGCCATT TTTTATCTC TGGCGCACGT TCTAAAATTT CAGCAATTTT CTCATTGGTT	25560
AAAAGCGGTG CTTCTTTCAG TTCGTGCTTA TCAATCAGCT GGAAATTTCT CTCCGCACGT	25620
GGCCGTAGTT GGGCACGAAC TTTTGAAAAC TTCACAACGT CGTCTGTAAT CGTCCATTCT	25680
CCTGTACCTT CCCAAGCCTG CACTGCTCGA GGCGCCACGT AATTATCAGC CCAGTATAAT	25740
AATTCTTCTT TTTCAATTTT AAAGGTCGAA ATATTGTCTA AACGAGGTTG AATAATCGTC	25800
ATTGCGACGG TTTCAAATTC GTAAATAATG TCGTACTTAT CCACCGCACC AAGCGCATAT	25860
AACATCAGTT GAGGGTTTAA ATACGCATCA ACAGGAACAC CTTTGCCGTA TTTCAAGTCG	25920
ATAATTTCAA TCGTCTTATC TGATAAGACA ACCACGTCCG AAGTTCCAAA TCCTTCTGGG	25980
ACCCATTTTG AAAAATCTAC TTTTGTCTT AATTGACTG TGGCATTTC GTATTGATTG	26040
ACACGTTCTT CTAATAATC ACAATAAGCG GTGACGTATT CTTTCATGGA TTCATCACAA	26100
TACGGATGGT CTTGtTCAA AAATTTCAAT CGTGAGTTA CCGCTCTAGC GGTTAATAGC	26160
TTAAACCGct TCGCCAAATA CAACTCTGCT AATTCGTGTG CAGTTGTTC CTCCTCCGCA	26220
TAAGAGcTGC CGCGATCTTT TACTTTTTCC TCTAGTCGTG CCAAAGGTGG GCATATTAGC	26280
CACCGATGCG CACTACTCGC CCCTAATAAG GCATGATTC CTAATGGCAT GCTTATTCAC	26340
CTACCAACGT TTCTAAGTCA GTGATAAATT TACCGTAATG CTCTTCTTTC AAGTCTGAAA	26400
GTTTTTCTGC GTTGTAACGA CTAAAGCAGG TCTTTATCAT AGGTCGCTTG CCTTTATCCA	26460
TAGCCTTTCT CATCGCTTCT TGCACATCTG CTTTGTGTG CCCTAGGTGT AAATCGGCTG	26520
TTGCTGATAC TTCTTCTGTT TCTGAATCAG TAGGCGCGAC TTTCTCCCA GTTTTCGATG	26580
GTTGCGGTGC TGAGGTTGAG GCCGTAGAAA CTTCTTCTTT CTTCTTCGCA GTTGCTTTTT	26640
TCTTTTTTGC AGGTtCTTTC TTCGTGCTT TTTCTACTTC TTCTGTAGCT TTTTATCCG	26700
CTGTATTTAA AGATTGCGTG ACTTCTACAA CGTTGTCTGT TGCTAACTGC GATAAAATTT	26760
CTTTTAACTC TTGTGCTGAA TCTGCTTCAA TATTTAAATA GATGTTTGCC ATTTTTTATT	26820
CCTCCAGTTA TTTTAATTGG TTAATGTCTT CATCTGTTAC ATTGAAAATC GCTTTAAAT	26880
TTGCGTGATA ATGCAACGGT GGGTCTACAT TACGTTTTTC ATAAACTTTT ATCAGCTTCA	26940
CAGGAAGTAG CATTCTTTTT GCTAATCTT CTTGTGACAT GCCCGCTTTT CGTCTAATTT	27000
TCTTGACGG CGATAACATG AGTTGCTACA CTTCTATCT GGTTAGCTC TAGGGTTAAA	27060
CATTCCCATG CTTCAATGC TTGGATTCCC TGTTTTCTAA GGCAATTAAC AAGCGTTGTT	27120
TTTCCTGTAG GCCCtGCTG ACCTTTCACA ATGATTGGTT TTCCCATT ATATGCTTTT	27180
AATATCTGCT CTTTTTGCTT CGATGTTAAA AAGTCAGGAA GTACTATCTT TTTATCGCTC	27240
ATCAACTTTC ACCGCCAATT CCCGATAAAA TCCAGGTACT TCTTCTTTTA GTTCTGTTT	27300

TGTGAAACGT TGTTTTAGCT TGTCTGATTT TGCTGTACAA ACAAACGAT TCTTTTCAGT	27360
GTTGTAAGAT AAATATTGCT GTTTCCCGT AGCATTTATC GCGTATGGCA AAGGAATTC	27420
ATACATTGGT TCTTTTTTCA CTGTGTAACC ATCAAGAATC GCACGTGTAT GCAGTTCTCG	27480
ATTTTCTTCT ACTTCTTTCG TCCACTCAAG GCCTTCCCG GTAAGCAACT CTTCATCTGT	27540
AATAGGATCT TCTAACCAAT GATCCCATCC CATCTTATTA ATTTGCCAA TACACCAACT	27600
CTTATAATCA ATATGTGCTT CTTTACTTTT CTGTACCCAT TTATCAAATT CTTTGTAGAC	27660
TTCAATTGAT TTTTATAGTG AAGTTGCTGT CCATTCAATA AAATCAGTAA TTCCTAATAT	27720
ATGCTCTAGA AAAGGAGCAC TAATAATTGA CAATTCTCTA TTCGCTTTAT TTAATTCAAT	27780
AACTGTTTTT TCTTTATGCC ATTTCCATAA ATCGGTTTTT GTAGGCTCAA GTCCTGTAGC	27840
CCAGACATAC CCTTGTCTT CCAAAAATGC CATTAAAGTTA TCATATGCTT CTTGTGTTTT	27900
TACGTGATAA TAGTTTCTCA TTGCGTTTC TCCTTTTTCG TGCTATTATT TACTTGTATA	27960
TTTTTTGTTT AGTAGCTTAC TTCGTGGCG GACGAGGTAG GCTCTTTTGT TTTCAATACA	28020
GCAATCATCA CCTAACACCC CACTTCGTTG AGCATTTTTT GATAAGCAAC TTCGTAACGT	28080
TCTAACTCTG CTTGAAAATG CTTTAGCGTG CGGAGGTCCT GCATTGTCGG ATGCTTTGCA	28140
CTTTCGTGAC GTACTGCATC TCTTAACGTT TCAATCTTTT CTCGTACTGC TTCACGTACT	28200
AAAAAAGCTT CATTAGCTGT TAACATTTAT TTCACCTCAC GATCTTCAAG CGCTAAATCA	28260
TAAATTAAAA GCCAAATGAT AAAAACCCT ATATAAATGT TTTGGATTAA TGGTCCAATA	28320
TTGCCACCTA CCAATAGACC TAGTCCGAAT ACAACCAACA GTGCCGTAT ACGTCTTAAT	28380
TGATATATTT TTTTCACTGT AATCATCCTT TCTTTCCAAT GTAAACATCT GCAAATCTTT	28440
CATCGTACAC CTCATAAATT GAGATATTTT TTTCTGGATC TGTGCAGCAA TAACGTAAAT	28500
TAACTCTTCT TTCAAATCTT TTAAATTGAG CGGCATCTAA AAGTTTGTAG TCAATTGATC	28560
CATAAAATGT TTCATAAAT TCCACGCTAG CCTTTTGCAA TATTTTGTGT AATTTTTTCA	28620
TTGCTTTCAC CTCCCTAAAA CCATCACAGA ACCGTAAAA TTTCCGATG CTTTGTCTC	28680
TGGCAAACGT TCATCGCCGC CTACAATTC ATTTGTTTGA TGGTTATACA CTCGTTTTTG	28740
CCCTTTAAAC AACACAACG AAATGTTGTC GTATGTGCTA CAAGTGCTT GCACTTTTAA	28800
ATCTTCGCCT TTGTAGCAA TAATCATTTT TGTAACCT CCTACGGAAT TAATCTGTGC	28860
ATGCGGTTCT TATCTTCTAA ACGACTTAAA AAATCTAATT CGTTTTGAAT TTTCCGCTT	28920
CTCTGTCTGT TAATAAATCA GCGCTTCTAA GAGCTGCCCT ATCTGATTGT TTTGCTTGC	28980
GATCTTTTTG AATTTGGCAT AATATCCACG CTTCTTGTTC TGTCGTATAA GCCACGTACT	29040
ATCCGCCCTT TCTTTTCTTG TAACGGTTTT TATCCATCCA GCGAACAAAC TCATCAAACA	29100
CATCTAACTC AATTAATACA ATCTTATGTG TTGGATGAAT TACACCAGAC GAAATTGCG	29160
GCGTGTCTG CATTAAAGCC AGCCATTTCT CTAACTAGA TTTAGTGATT GAGTACATTT	29220
TTGTACTAC GGCTTTATTT CCATATTTAA TTGACGCACG ATCTGGATTC TGTTGACAG	29280

CTTCCAGATC AACCTTAACC GGTTTTGGCA TACCCCTTAC CCCCTCTCAT GCTTTAGCTC	29340
GTTTAGTTCA ACTAGCACAG CCACATAATC GGTAACAAA GAATAACGGT CCTGATGTAG	29400
TTTTTCTGTC CATCGATAAG ACACATCTAA CTGCAGTAGT TGATGAAGTA AAAAATTTT	29460
TTACGAATT GATGCATCTA GATTTTCTTG TGCTTTCCAA TAACGATTG TTCCATCGCA	29520
TAATGAGCTT GATGAAATTT TTTCTCTGCA TCGATTAATT CTGTTGTCG TTTTAAAAAC	29580
TTCTTTAAAT AGATCTTTGA CAATTAAATG CAGTCTTGAT TTAACAATGT CTTCTGCCGT	29640
TAGTGCGGGC ATTCGTGTG CATAGTTCAT TTTCTAAGCT CCTTTTGTTA AAAAAATAGT	29700
AATTACCACT TCGTAAATA AGAATTACGT TTTATAATAA AAAAGAATT AAGAATTACG	29760
AAAATCGTAT TCCACAAAG TTTCAAAATC TACATTAAGT AATTTTGCTA ACATAATTGA	29820
TAGTTTGAAA CTCGGGCAGG AACGGCCATT CTCTATATTA CGAATAGTTG TTTCAGTAAC	29880
ACCAGTTAAC TCAGCTAAAC GTTTTGTAGT TAATGACTTT TTGATTCTCG CTGCTCGAA	29940
AGACCCTCTT GTACAATTCA TATGAACATC TCCTCTCATT TAATTTACGT AATTGCCAAT	30000
TACATAAATG AGTATACAAT GTAATTAAG ATTACGCAAG AGGTTTATTT GCTATTTTAA	30060
CAAAAATAAC ATAAAGGAGC TTTAAATGA CCTCATTTGC GCTGAGATTA AAGGAATTAA	30120
GAAAAAATAA AAAAATGTCA CAGCAAGAAC TTGCAACTTT TTTAGGTATG AAAAGAGAAA	30180
ATATTTCTAA CTATGAGCGA GGGGTAATTA CCAATGTATC TAGTGATACA TTAGAAAAGA	30240
TTTCAGAACT TTTTGATGTA TCTATAGACT ATTTGCTAGG AAAATCAGAT ACATCTAATC	30300
AGAAAAATGA ATCGATGGCT AAACAAGTAA TGATGCGCAT GGATACAGAT GGACTTTCTT	30360
CAAAGCAAGT AAAAGAATA GAGGATGAGA TGGAACGTTT CTTTTCATGG AGAATAGAAG	30420
AAATAAAAAA AGAAAATTTA AGTGAATAAT TACAAGGTGG TTTTCAGGAT GACAGTAGAA	30480
GTAGACGAGT ACCTTAACTT TTGCGGTACC ATTAACGAAT TTATTTACGC ACACATGTAA	30540
TGTTTGGGAA TGAGTGTTAA TAATTATGAG CATAGATATA TTTGGGATGA AATATTAACA	30600
TCTAAATCAA TTAAAATAAG GCCCTTTCCA TTTGAAAAA CAGCACGTAG ATCTATATCA	30660
GGTATGATTA TTAAAGATGA TTACGAGACT ACGCTAGCTT ATAATTCTAA CATGGGCGAA	30720
AAAAGAAAAA ATTTACAGAT ATCTCATGAA CTTATTCACG CCATGTATCA TTTAGATAGC	30780
GAAATAAGG TATTCACCGA CACAAAAGAT ACTTTATCCT ATTCACTGGC AGATATCTTA	30840
CCAGAATTC AAGCGAACAT TGGAGCATCT TCTATACTAC TACCCGAACC TGTCTTATT	30900
AATGAGCTGA AAAAAGGAAC TCCTCCTTAT TTTATCTCAA ACCGTTACGG TATATCTGAA	30960
CAAGCTATTT TTATGAGATT ACTTCAGCAA ATGCAGGCTA GTTTTGAAGC ATCTTACGTT	31020
GCCGCCTATG ATACGGCCAA TAAGATAATG AACGGCAACT CTAAAAATTT GGCAATTGAA	31080
TTAGGACGAA ACTTAGAAAG AAAAATTTTA TACAGCAACC CTTTTTATGA AGCAATTACG	31140
CTATAATGAA TTTATATTAA ATCGAAAGAA GGATGATAGA AATGAAAAA ATGATTAAAT	31200
TTGCAGGCAT TGCTCTTATT TTTGCAGCTC TTCTCTCTGC CTGTAGCAAC GCAAAAAATA	31260

ATACACAAAA GAAAGCCGAA ACTGCTGCCC AGTCAAGCAC TATTGAAGCT TCAGACAGTA	31320
ACGAAAACGA GCCTAATACA GAAAACATAA CCCAAGCAGT TAAACAGTTA GAAGAAAAAT	31380
TTAACTCTGA CGAGAAATTA GTAAAAATAG ATGTTAAAAA TAATGTTAAA GATGACACAT	31440
CAGATAACCC TCACGCTGTC ATTACGGTTA AGGTAATTAA TGATGAAGCA AAAAAAATA	31500
TGGAAGAAAT GCAGACTGCG ATAGATTCCA ACTCAGGTAC AGAGGCACAA AAGACTGCCA	31560
TATACGGAAT TCAATTAAAT GTTGAAGAAG TAGCCAAAAC ATTAGAAAAT GATAACGATG	31620
TTATTTCTTT CATCACACCT TACACGAATG GGAACGACAG AACCATAGCA AAATCAACTA	31680
AAAAAGAAA TATTATTCCG TTAGTAAAAT AAAAAAGTGC ATCCCCGCC GGCAAGCTAG	31740
TGGATGCACC TTATCAGAAA ATAAACCAAA GGCTTATTTA ATGCTATTAT ATCAAAAAAT	31800
AAGCCTCATT ACGAGGTTTT TATACACCTC CAAAAGAACG TAAGTTCCCA TAAAGGAGA	31860
GTTGAATATG TGGGTAGAAA AGAAAAAGGA TAAAAAGGA AATATCACAT ACCAATATCG	31920
CGAAAGATAC ACCGACCCTC GGACCGGTAA AACACAAAAA GTCACCACGA CTTTAGCAAA	31980
AAATACAAAA CAAGCTGAGA aTGCAGCAGC TCGTGAGTTA GAAATAAAAA TCCAAAAGAT	32040
ACTGACCGAC TACGAAGAAG AGAAACGTCT AGAAAATTTA ATCAAAAACG AGGGCGTAAT	32100
ACCAAATGAT AAAACGTTAG AAGATATGAT AAAAGAATGG TTTGATTATA TTCAGGATCC	32160
AAAATCAAAA GAACGTAAAA AAGGTTCCAC GCTCTCGGGA TATAGCTATG GCATTGATTC	32220
ACTCTTAAAT GAGTTCTTAG TAATAGAAAA AGATAAACGA ATCCAGGATT TAACTTTTGA	32280
TGATATACAA GCATTTTATG ACCGTATGAT TTTTGACTTT GAATATCAAA AGTCCATAT	32340
AAAAAGATTC CGTGCGATTA TCCGTGGTGC ATTTACTCTC GCCTATAAGA AAAGATACAT	32400
TAAAAATCTT GATGCCATTA ATCTTTCTAA AATAGTAACG CCAGCAAAAA CGGTGGaTGA	32460
TCTTATTAAT GAGCAAGTAC CTAGATACTT AGAAAAAGCC GAGGCTGAAC AAATATTTCA	32520
AGTTTTAAAT GGATACAACC CACTTTATAC GCAAATACTC GAACTACAAT TTCACACTGG	32580
GATGCGTTTT GGTGAATTGA TTGCTTTAGA ATCAAAAAAC TTTCAGGACG GATTTTTGTT	32640
AGATATTCAT GGAACCTACG ATCATGCAAC ACGTAGTCAC ACCCGAGGAC AAAAAGTCCC	32700
GCCAAAAACA AGAAAATCAT TTCGTACTAT TGAGTTAGGT GATGCTGCAG TACAGATTAT	32760
AAAAGACC	32768

## (2) INFORMATION FOR SEQ ID NO: 101:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7964 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

AGTGAGCGGG TCACTCTAGA GAAAAGGAGT AGAAAAAAT GGCGGATTTA AAACGACTGT

662

TCATTGGGAA ACCATTGAAA TCAGCCGAAA ATGATGAACA TAAATTAAGC AGATTTCAG	120
CGCTTGCGTT GTTATCTTCT GATGCACTTT CATCCATTGC GTATGGTACA GAACAAATAG	180
TGGTGGTTCT GGTACTTTG TCGGCGGCTG CGATTTGGTA TTCATTGCCA ATTGCTGCAT	240
TTGTTATTAT TTTACTTATT TCTTTAACAT TGTCGTACCG ACAAATTATT CATGCGTATC	300
CGCATGGCGG CGGCGCATAC GTTGTTAGTA GTGAAAACCT CGGCAGAAAT GCAGGTTTAT	360
TGGCTGGTGG TTCGTTGTTA GTCGATTATA TGCTAACAGT AGCCGTTTCA GTTTCAGCGG	420
GTGCAGAGGC GATTATATCG GCGATTCCAG CGTTATATGG TCATCAAGTA GCAATTTCTA	480
TTGGGATTGT GATATTGATT ACCTTAATGA ACTTGAGGGG CCTGAGAGAA TCAGCTTCGT	540
TCTTAATGTT GCCAGTGAT AGTTTTATCG CAATTATTAC CTTACTGATT GTTGTGGGC	600
TATTTAAAT TGTGACTGGC GCACAGCCAT TGAATGCGAC AgcTTTACCA GGAGCCGTGG	660
TTCCAGGAAT TTCTATTGCC TTGGTTTTAC GGGCTTTTTT ATCTGGTTCT TCTTCCTTAA	720
CAGGGGTCGA AGCAATCAGT AATGCAGTGC CATTCTTTAA GAAACCACGA GCAAAAAATG	780
CTGCTGGGAC GTTAGCGTTA ATGGCAACGA TTTTAGGGTT CTTCTTCGTG GGGATTACGT	840
TTATTAAC TA TGGTATGGG ATTGTGCCAA AAGAAGAAGT GACAGTTCTT TCACAAATTG	900
GGAAAGCGGT CTTTGGTCAA AATATTTTAT ATTATTTATT ACAGTTTGC ACAGCCTTAA	960
TTTTAGCGGT TGCCGCTAAC ACAGGATTCT CAGCCTTTCC AGTTTGGCT TATAACTTAG	1020
CCAAAGATAA GTTTATGCCG CATATGTATA TGGATCGTGG CGACCGTTTG GGCTATTC AA	1080
ATGGTATTTT AACTTTAGCT GCTGGTTCCA TTGTTTTATT ATTGATTTTC CAAGGTCCA	1140
CAGAACGTTT AATCCCATTG TATTCCATTG GTGTGTTTAT TCCGTTTGCT TTGTCTCAAT	1200
cGGGGATGGT TGTA AAaTGG CGAAAAGAAA CTA AAAATTG GTTGCCAAA TCAATTGCTA	1260
ATATTGTCGG CGCCTTTATC TCTTCGCGA TTATCGCCAT TTTGTTTATT TATCGATTAG	1320
GGGACATCTG GCCGTTCTTT ATTATTATGC CAGTGTTGAT TTATGCATTT TATCGGGTCA	1380
ATACTCACTA TAAAAATGTG GCGGAACAGC TACGTTTAGA AGATGGGGCA CAGTTACATG	1440
AGTTTGACGG GAATACAGTT ATTGTTTTAG TAGGGAACTG AACCAAGGCA AATGTCGGCG	1500
CCTTGAATTA TGCGCGTTCC ATTGGCGATT ATGTCGTTGC CATGCATGTC TCAATGGATG	1560
AAAATGTTGA GAAGGAAAAA GAAATTCAGG AAGAGTTTAA AAAACACTTC CCGGACGTTT	1620
GTCTATCAAT TGTCATTCT TCGTATCGTT CACTGCAAAA TCCTATCTTA CGTTATGTTG	1680
ATTTGGTTAG TAAAAATGCC ACGAAGCATA ACTACAGTAC AACGGTTTTA GTGCCACAAT	1740
TTGTACCAA TAAACGTTGG CAAAATATTT TGCATAACCA AACAAGTTTA CGCTTGCGGA	1800
TTCGTCTAGC GTGGCGGGAA AATATCATCG TGGCTACCTA TAGTTATCAC TTGAAAAAAT	1860
AAATAAAAAA ATCTGCCATC TTTTGATGGC AGATTTTTTT ATTGACGTAA TTGTTTTTGA	1920
TAAATTCTAG CGCAAGACTG CTCTTCGGTG GTTACCATTG TCAAAAATTC CCAGTCATAT	1980
CTTTCATAAA AAGTAGTATG GTCTGTGAGC AAATAAAGCT TTTTAAATCC AGAATTAGCT	2040

AGAAAATCAC CAACTGTTGT TAAAAGAATT CCGCTAAGC CTTGGTGACG AAATGTTGGT	2100
TCGACATACA ATGCACAAAG GTTTGGCGTA AGGTCTGCAC GAFTGTGGAA GTCATTATCA	2160
ATTACGCCCA CGCCTGCAAC GATGTCCTGC TGTTTATTCA AGCAGACCCA CCAATGTGGT	2220
ACGTGGTTTT CGCTAGTTAG CATTTCCTGC ATACTTGCTT GATACGCTTC AGCGGGGATT	2280
TGCCACTTTT GGCTAAACCA TTGAGCAGCT GACGGAATTA AGTGTGGTTG TTGGTTTAAA	2340
TCGATAATAG TATAGTTAGA CATTATTGAT CTCCTAGATA AAAAAACAGA TCCCATTGAA	2400
AAGAAACAAT GTTGATCTG CTCTGTATTT TTTATTAATT GGAATGTTG CTAGCTGGGT	2460
CTAAACGCCG TTCAATAATC CCTAGAACCC AATCTGTGAT AATCGCCATA AAGGCCGTTG	2520
GTAAGGCACC GACTAAAATG ATAGAAGCAC CATCTGTGGC ATTGGTTCCA CGAATAATAA	2580
TATCCCCTAA ACCACCAGCA CCAACGAAAG CACCAATCGC AGTAATCCCG ATTGCGACGA	2640
CTAACGCATT TCGAATACCT GCCATGATTA CTGAAACAGA TAAAGGTAAC TCGACCATAT	2700
AAAGAAGTTG TACTTTTGTC ATACCCATTC CTTTACCAAC ATCTAAAATA TTGCGATCCA	2760
CTTGAATCAT GCCAGTGAC GTATTCTTGA TGATTGGTAG GAGAGAATAA AGGAAGACCG	2820
TTACAACCAC CGTGTTGACG CCTAAGCCCA TTCCCAACAT TAAATCGAA AGCATTGCCA	2880
AAGCTGGTAC TGTTTGGATA ATGTTGCTA AACGGATAAT CCAGTTGGCT AAGCTACGTT	2940
TACGTGAAAT CATAATCCCA ATAGGaaTCC CGaCAATTGC AGCAAAGATT ACGCCATAAA	3000
TCGAAATTaA AAAGTGACGT AAAAATTGTT GaAAGACGTA GCCGCCATTT tGTTGGaAAT	3060
AGTAGATAAA TTGTTGGAAT AAATcATATT ATTCatCTGk TACTCCCCCT CAAAATAGTT	3120
ATTTtCTTT AAGAAaCGTT CGGCTACGAC AGCAGGTTCA ATTAAATCAT TATCTGCTTG	3180
ATAGTTCAAT TCTTGCAATT TTTTGGTAGA AATTTTACCA TCTAAGCGAT TTAAGACATC	3240
TTTAATTTC GCGGTTTCTT TTAATAATTT GTCACTAACC ACAGGAGCGG CATCATAAGG	3300
CGGGAAGAAG CGCTTGTCAT CTTTTAACAT GACTAAATCG TAACTACCAA TCCGGCCATC	3360
TGTTGAGTAA CCTAAGATGA CGTCCATTTT TCCAGCGTTG ACTGCATCGT ACACCAAGCC	3420
AATTTGCATT GGTAAATAG ATTTAAATTG GAAGTTGTAT TCCTTTTTAA AGCCTTCATA	3480
ACCGTCGCCT TCACGTGTGA TCCACGCAGT ATCGACACCA GCAACTAATT TGTCTGCGAC	3540
TTTCCCTAAA TCACTGACTG TTTTAAATT GTATTTCTCA GCCGTTTCTT TTGTAATAA	3600
GAATACATAA GTATTTTCAA AACCATATGA AGGGAACCAA GTTTGTTGAT AACGTTTTTC	3660
AAATTCTGAT TGAAC TAGAT TGAACGCTTT TTTAGGGTCT TTTCCGCTG GTAAGCCAAG	3720
CGTAGTAGTC ACGTCAGTAC CAGTATAACG AGTGCGAGAA ATTGAGGCAT CTCCGTTTAA	3780
CATTGCTTGG TGATTGATGG TAGTAGTGGC TAAGTTGTTA ATCACATTTG TTTTTTATC	3840
TGTGTAATGC TCCACCATAT CCGCTACAAT GCTCCCTAAA ATCTGCGCTT CTGAAGTAAT	3900
CCCACCTGTA ATGGCAACAG TGTTATCATC AGAAGTTGAA GCGAGACCAG GTAAAGAACA	3960
ACCTGCCAAT AAAAGTAAAC TGCTTAAGAA GAGTAAGCTT AACTTGAGTT TCTTTTTCAT	4020

CATTATTCGC	CCTCCCTTAA	AGCTTTTGGC	GTTAAACGAT	TTTCTAATAA	TCCAAGAAGA	4080
ACATCTGCTA	AAAGCGCTAA	AATAGTTACC	GGAATAGTCC	CGCCAAAGAT	TAATTCTGGC	4140
TGATAGTTAT	TTAATCCACT	GAAAATCATA	TCACCAAGAC	CACCTGCACC	AATATAAGAA	4200
GCCAGTGTAG	CCCAAGCAAT	GACATAGACA	GCAGCTAAAC	GAATGCCGGC	CATAATTGTT	4260
GGCATGGCAA	TGGGTAATTC	AACACTAAAA	ATAGACTGTA	AATTGGTCAT	GCCCATACCT	4320
TTAGCAACAT	CTCGGTAATT	TTCATCGACA	TTTTTCATTC	CGATATACGT	GTTTCTTAAA	4380
ATAGGTAGAA	GGGAATAGAT	GAATAGCGCG	ATAATAGCGG	GTGTTTTGCC	GACACCAAAA	4440
ATTGGAATCA	TCAAAGCTAA	TAAAGCCAAC	GAAGGAACCG	TTTGTAAGGC	ACTTGTTAAA	4500
CCAATGACAA	TATTGGCTAC	TTTGGGCAAA	CGTGTTAACC	AAACGCCAAC	GGGCACAGCG	4560
ATAGCGACCC	CTAATAAAAG	TGCAAAGAAT	GAAATATAAA	TATGTTCAAT	ACTTTTTGAA	4620
ATCATTCTTG	ATCCATATTC	TTGAAAGAAG	CTACTCATTG	yTaCGCCTCC	TCTTTATCAG	4680
TTTCGGACTC	GCTTTGTTTT	GCTTCTACAG	CTTCACTAAT	CGTTGTTTCA	TCGCCCCAGA	4740
TAACATCATA	AACAATGTCA	ACAAGGGAAG	CACGTGTAA	GATGCCACC	ACTCGTTTTT	4800
GTTCATCAAC	GACTGGCACA	TATTTTAAGC	CGCGTTTTAA	GATTCTTTGT	AAAGCATCAC	4860
GTAAAGAGC	GGTTTTTTGA	ACAAAGAAGA	CATCTTTATT	CAAGATATCA	CCGACACTAG	4920
AAGCTTTGCC	ACGTTGCTGA	TCTAATGTTT	CAACATCGAT	AAAGCCTTTA	AGCACATGGG	4980
AATTATCAAC	AACTAATAAG	GTATCAACCC	GTTTTTCACG	CATCAATTTA	ATCGCTTCTT	5040
GTAAAGATTT	TTCCGGCGTG	ATAGTAATAG	CTGAATTTAA	CATGACTTCA	TCAACAGTGG	5100
TGAAATCAGG	TTTGGCTTGT	AATAGGCGAT	CTTCTCCGAT	GAGTTCTTCT	ACAAATTCAT	5160
TGGCTGGATG	ACGTAAGATA	TTATCTGGTG	TGTCAAATTG	AATTACTTTG	CCTTCGCTCA	5220
TAATGGCAAT	TTTATTTGCT	AATTTCAACG	CTTCATCCAT	ATCATGGGTA	ACAAAGACAA	5280
TAGTTTTTCC	TAAACGTTCT	TGCAAATCTT	TCACTAAATC	TTGTAATGAA	TCACGGGTGA	5340
TAGGATCTAA	AGCACCAAAA	GGTTCATCCA	TTAGAATAAT	ATCTTGATTG	GCAGCTAAAG	5400
CCCGCACAAAC	GCCGATTCTT	TGTTGTTGTC	CGCCAGAAAG	TTCATTGGGA	TAGCGGTCCA	5460
ACATTTCTCG	TGGTAACTCA	ACCAAATCAA	TCATTTTTTC	GGCTATCTTA	TTGCGTTCTT	5520
CCACAGGTAC	TTTTAATAAT	TTAGGCACTA	AAACGATATT	TTCACGAATG	GTCAATGTGAG	5580
GCATCAAGCC	GATATTTTGA	ATGACATAGC	CAATTTGGCG	ACGTAGCTCT	ACAGGATTGA	5640
TTTTTTGAAT	ATCTkGGCCa	TCaATTAAAA	TTTTTCCTTT	TGAAGGGTCG	GTCAATGCGGT	5700
TAAGCATCCG	CATTGATGTT	GTTTTTCCAC	TACCACTTGT	TCCGATGAAG	CAGATAAATT	5760
CACCTTTATC	GAAAGACAAA	TTAATGTCAT	CAACGGCGAC	TTTCCGCCT	TTATAAAATT	5820
TTGATACATG	CTGAAATTCT	ATCAACTTAC	TCACCTCAAC	TAAATATTTT	GAAAATGTAT	5880
CTTATTACTA	GTATGCTATA	AAAATGCGAT	TTGACAAAAT	GTAATCTATT	TTTAATGGGA	5940
AAAAGAAGGC	GGTAAAATTT	TTTAAATTCG	TTCTAAAAAA	AAGAAAAGAT	AACTATTGTT	6000



TAGTAAAAGT TATCTGAACG AAACAAGATG TTAAAAAGAG TATTCTGAGA AGGAAGAAAA	6060
GCCGCTATAA CAAAGAAAAG TGTCTTTTGT AGCAAGTTGA CGGGAGCTGA GTTTTTCATT	6120
AGAATGGAAA GAGAACAAGG GAATAAAATA AAAAAAGGCG CTTCATGATT TAGCGGCAAC	6180
TAAATCACGA AGCCTTGAAT AGTCACCTGT GCAAGTGGAC CATCAAGTTG CCCGTTCAAT	6240
CAATGGAAAC CATTGATCAC TTTTATTTTA CTGAATTTTT GAAAAATTA CCAGTTTTTT	6300
TAAGAAGTAA GAAGTGGAAAG AACGAAGGAT GCTTTTTATG GAATGTGCAA TGACGTTTAG	6360
AAAAATCTGA AGGTCATTTT TTATTCTCTT GTTCAAACAT TCAAACATAG GGAAGAGAG	6420
ACGGACGTAA TTTTGTCAAC GATTATTACG TCCGTCTTCC GTTGTCTTCT AAAGCGTTGG	6480
TTACATGTCT TTGTAGAGGT TGGCTTCAAT TGCAGTCCGT TTTGGTTCTA AAAAAGGCGG	6540
TAAAGCCAAC GTTTCCTTA AATGTTCCAG AGGCTCATCC AATGTAAAAC CAGGAGTCAA	6600
AGTAGCAAAT TCAACCCGAT TCCCGCCTGG TTCACGAATG TATAAAGTTT TAAAATATCC	6660
ACGACTTACA TACATTTCAA TTTCCAGCC TTGTGCTAAA GCGCGCGTGT GTAGTTCTTG	6720
TAATGCTTCA TCATCAGCCA CAGAAAAGC AATATGATCC ATACTGCCAC GCCCATTCG	6780
GCTTGTTTTA TCCGACGTTG TTTGATGAAC AGAAAAGCTT TCCGTCTCAC TTAAAATAAT	6840
CGTATTGCCA ACTGCTTGCC AACCAAGAAA TGTTTCAAAA AAGGCCAAGG TTTTTCAGA	6900
ATCTGGCACA TGAAATTCAG TAGAGGCCAA ACCGATAATT TGAACCGCTG CTGGAATGTC	6960
GTTAGAAACA ACGAGTGTTC CTTTAAATGG ATTTTCAGGA ACTTCTGTCA GGATAATATG	7020
CGTATCATCT GGATCACTGA AATGTAGGCT CTTTCTTTT TTGTAAATG GAATTTTGTT	7080
TTCTGTAAA CGTTGTTCCC AAAAGGAAAG GCTGTTTTGG GGAATTTTTA AGCCAATCGT	7140
GGCTAAATAA TTTTTTCTT CATAGCGGTT GCCTAATAAG GGCACGATAA AAAAAGTTAC	7200
TAAGCTTCCC GCGTGCCTT GATAATCGCC ATAAAAATAA TGAGGCATGC GATTGTTGTC	7260
TTGATTTACG GTATTTTGA CAAAACGTAA GCCTAACAAT TTTGTATAGA ATGATTGGTT	7320
TTCTTGACCA TGACGCGCTA ATAGTGAAAC ATGATGAAAT TGAGCATGCA TTTTGAGTTC	7380
CTCCTCTAAA TAACTATGTA TACTTACTTA AAGTAAGTAT AGTGCTTCTG ACATTTTTTG	7440
ACAACATTTT AGTTTGATTC AGAAGGCTGA CATAGCCAGA AACmAGAAAA AGTGTTAGAA	7500
TAACCAAGAT wAAAACTAGT AGAAAGAGTG ACGGAAATGC TAAGTACaGA AGATTTTGAT	7560
TTTGATTTAC CAGAAGAATT AATTGCTCAA ACCCCACTAA AAGATmGCGC TAGTTCACGA	7620
TTATTAGTTG TAAATAAAGA AaCAGGGGAC ATGGAGGATA AACATTtCCA TGATATTCTT	7680
GATGAATTAC AGCCTGGGGA TGCATTAGTC ATGAATAATA CACGGGTACT GCCGGCCCGT	7740
TTATATGGAG AAAAACCTGA AACAGGCGGT CATTTGGAAG TTTTACTTTT AACAAATACA	7800
GAAGGCGATA CATGGGAAAC CTTGATTAAA CCCGCGAAAC GTGCAAAAGT CGGCACGGAA	7860
ATTCAATTTG GCGATGGTCG CCTAAAAGCG GTGGTTAAAG AAGAGTTGGA ACATGGTGGT	7920
CGAATCATTG AATTTAAATA TGATGGTATT TTCTTAGAAA TTTT	7964

## (2) INFORMATION FOR SEQ ID NO: 102:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7215 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

```

AGAAGTGAAG AGAGCTTTTG ACTTACGCTn nTATTATGAG GAACGAACCT TTCAGGCTGG      60
GAAATAAGG AGGTATCAAA AGAATGAATT TGTTCAAaAA GAACACAATCA AAGAACACAGA      120
CGTTATCTAG TAACGTTGAA AAAGTAGATA AAAAAGTAA AAAACGTGTA CAGCAACAAG      180
CAAATCCTAC CGTACAGGAG AGTATTTTTT ATACGAGTCA GTTTGAAGAA GAGCTCATGC      240
ACATTGTAGA GGATGAGTTT TCTAAATGCT ATAAACTtGG AGAAGTGGAC TACGAAATTG      300
CTGTAGAAGA AGATCmATTA AACaCGGTAA TGGGGTATGC AGAAGGGTTA AATACACTAG      360
ACAAAAAATC TCGGTATCaG CTGCTGGTTG TTAATAAAAG AGAAGAAGAG TCGTTGATGG      420
AAGACGTTTT ATTGGCGTAT CaGGGAGATC ACTTAGATAA cTACCGTCAA GAAATCAATG      480
AAATTATAAC GAAGCAGTAT GAACAAGACG AACGAACTT TAAGATTGAA AAATATGCTA      540
TTTTACCAC AAAAGTTTCG TCCATGAAAC AAGCTAACAA AGAGTTAGAA GGAATCATGC      600
AGAAGTTTGC CAATCGTTTT GAAGCCAACC AAATTAAATT AGAGGTAACC CCAATGACTG      660
GCTTAGAGCG ATTGTCTTTG TTCAGCTATT TTTGCGCCA CGGACAATAT TTTGGCACA      720
GCTATCAAGA TATTGCCGTA TCAGGGCTAA CGTCTAAAGC GTTTATTGTT CCAAGTAAAA      780
TCAAGTTTCC AAACAATAAA GCTTATTTTC GGTAGGAGA AACTACGCA GCTGTTTAA      840
CCATTCAACA ATACCAAAA TACCTAGAAG ATAAGTTAAT CAAAGAATTA TGCGCTACAG      900
GAAGAGAGTT AGCGATTAGC ATTCATGCCA GACCTTATGA CATGATGGAA GCCAGAAAGG      960
CAATACAGGG AACCAAGACG TTAAATGATG TTGCCATTCA AAAGCAACAG AAGCAAACT      1020
TCAGAAGTGG TGTTTCAGAG GATGCGATTT CAGGAGAAGC AAAGGAAATC CAAGAAACAA      1080
CGAACGCGTT AATGGATGAA ATAAAAATA ATGGCCAAA GTTATTTAAT GGCCTATTTT      1140
CAGTTTTCTT AATTGCAGAA ACAGAAAAGG AACTTTTAGA AGCTGTTGAA TCGGTGAAAA      1200
ACGTAGGAGC AACCTGGCAA GTTGTGTTTG ATACGGTAGA CAATTACAA GAAGAAGCAC      1260
TCAATACTAT CTTACCGATT GGAAAACCTT ACTTAGACGT TGAAATGAGT TATATGCGtG      1320
ATCTAACmAC mACmAATATt GCTACACaG TACCTTTtAC CAATGTGGrA TTACAAgCCC      1380
AcAGGTCAGT TtATGGCCG GAATCmGTTG ACGAAGAaCA TGATtACTAT tGrCCGAAAA      1440
AAAGwTTAAT TwCACCAgTG GGCTTATTTc GGAcGcyGGT tCTGGTAAAG GATGGCGACC      1500
AAATGGGAAa TTCTTtCaC GCGTTtAAAA TtTCCGAAGa TCGCTTTTta ATTGTGATC      1560
CAGAAAGTGA ATACTTACCG ATTGGGAAAG AATTGGATGC AGAAATACTC GATATCTCaA      1620
CTGGTACGAA AAACCACTTG AATATTTTAG GAATGGTAGA CAAGCAGTTA TTGGATGAAG      1680

```

AGGACCAGGA ACTTGATTTA GTTAAAGAAA AATCTAATTT GTTATCCACT ATGTTTGAAT	1740
CGCTACTAAA AAGCTATACG GACGTGGATG CGGGACTTGT TGATCGTGTG ACGTTGTTGA	1800
CGTATGAACG CTTTGAAAAC AAAGAGACAG AGCCAACATT AGTCGATTGG TTTAGAATCT	1860
TAAAAGAACA AACGGATGAA GGTGCCAAAG AATTAGCGAA TAAAGTGGAA ACCTATTGTG	1920
TGGGTTCACA AGATATCTTT GCGCACCAAA CGAATATCGA CTTAACGGCA CCCTTTATTG	1980
TCTTTAATAC AAAAAAATTA GATGATCGCT TGaAAAAGTT TGCCATGAAA GTGATTTTAG	2040
ATCAGTTGTG GAAACAAGTA GTATCCGGAC AAGGAAAAGT AACGTCaCGT TTGTATTTTG	2100
ATGAATTACA AGTCAATTC ACAACGGaAG AAGAAGCAGA ATGGTTCTTA AATTATGGG	2160
CGCGTATTCG TAAGTATGGC GCCGTGACAA CAGGTATCAC GCAAAACCCT ACGACGTTAT	2220
TAGACAGTAA AGCGGGGCAA AAAATGATTG CCAATAGTGA ATTTTGTATT TTATTACGGC	2280
AAAAaCCAAT TGATTACAA CGATTAGTAG rAATTTTAAG TCTGACACCT GCATTgAAGA	2340
AGTATaTmA TGATCGAGCC CCACAAGGrA CGGGGcTAAT TtCTGCAGGT GGGGTAGTTG	2400
TTCCGTTTGA AAATCCTATT CCAAAAGAAA CACATCTATT TGAAATTATG AATACCGACG	2460
CTTAGGATGT GAGGACATGG TAGAAAGTAA GCGTACTTGG AAAGAAAAAG TACATGTTGA	2520
ATCTCAAGTA GAAGAGCGTC TTAGAAAAAA TTTGACCTTA AAATCAAGAG AAGCGACGAA	2580
tTGCAAAAAA AAGCATTAAA GACGACTAAA AAGGAGTTTA AACTTTCTAA GAAAAATTAC	2640
AAGCAATCAC TAAAAAACA TAACTTACC TTGAAAACAG GGAAGGAACC ATTAAGTATT	2700
AACCATTTAC TGGATAAAAA GAAATTTGTA GAGCAAAAGA AACGCAAAAA AGCACAGAAA	2760
ATTGCTTACA AACGAACGAA AAAAGTAGAT GAAACGCGTG TCGTTAACCA AGTGAAAAGG	2820
GAAACCAAGG AAGGGTTGAA AAGAGAAGCG ACGCAAAAAG TACGAACCAC GTTAACGCAG	2880
GAAGATACAC TGAATGAAGC AATGACCCTT TATGAAAAAA CACAACAAGC TAAATTTAAC	2940
ATGCGTACAG CTTTGAAAAC AGGAAAAACG GTTAAGAATC TTTCGGTGAA AACTGCAAAA	3000
GATACGTACG GATTGGGGAA TCGCTTATTT AACTTTTCAA GAGGACGTGG GTTTCAGCGC	3060
ACACCAAAAG ACTTTACGTT AAAAAACAA CTTATGAAAC AACTACGGAA TCGCGCTATG	3120
CGCTTTAAAG CAGCAAAAGA AGCGAAGAAA GCCGAACAAG GTATTGGCTT GATTGTTCT	3180
TTTTTTAATG GTCAGAAGAC ATTAGGGAAA GTCGCTGCAT TGATTCTAAA AAATCCAATT	3240
TCTTGGGTTG TTTTATTGGT TCTTTTCTT GTTTTTTTAT TGAGTGGCGT TGCAAGTAGT	3300
ACACAAAAAC CAGCGATTGT TCAAGAAGAG GAAGATTTAA CCGCTTCATG GACGTATTTT	3360
ACTAAATTAG ATGCGCAACA TACAGACGAT AACAACTGT TTTATAGCAA TATCGATGAT	3420
GTTCTTTTCT ACATGAATTA TCGGTACGAT GATTTTAAAT TATTAGATAT GGATTCCACT	3480
GGTACAAAGA ATTTTGAAAC GATTCTATCT GAACTTTGGA CGGCACTAAA CGGAAAGAAA	3540
CCTGATTATC AGTTGAAAAC AATGCAAAGT TTAGAGACAG ATAAGAAATC GAGTTATTTT	3600
ATCGAaGAGG AACAAAGCAA GCATTATCAA GAGaTAAAAA AAGAGTTAGG CTACCAAAACA	3660

TTGGATGATC TCCTTTCTTT TCCAGTAAAA ACTGACGCTT TAATTGTCaA TaAGCGTTAT	3720
GGCTATGACA AtCmAAAGAG aAGCTCACCT TATATAAAGg CATTGATGTA TTGATTGAAG	3780
ATAATCAGCC GTTTCATTCA CCAATGAaCG GCCAAaTCGT TTCAGTGCCT GATACaGAAA	3840
CGCTGGTTAT CGAAAAaGAG AAAGTAGCTA GGTTAACGAT TCGTGGTGTG AaTACCCTTC	3900
GACTAACAAA AGGTATrGAT GTTGAAGAAG GAACATTCT TGGGAATACA AAGAATTCAA	3960
CAGTTACGTT CCAATATGAA AAATATAAAA AAGAAACGAA AGACTGGTTT TTTGTAAATC	4020
CAGCCTTTTA TTTCCACGT GTAACCTACA CGCAAACGAC TCTTTTAGGA AGTGCAGAAT	4080
TTTCTCCTGG AGCAAGTGTC GAAAAAAGAG CGCAAGCTGT TTATGATTAT TTATCAAAAA	4140
AGGGATACAC CAAAGAGGGA ATTAGCGCGA TTTTAGGCAA TTTTTCGGTA GAAAGTGGCA	4200
TTAATCCAAA ACGTGCAGAA GGGGATTATT TGAATCCGCC TGTAGGTGCG CATGGGAATT	4260
CATGGGACGA ACCGAGTTGG CTTGCTATGG GAGGACCGCA AATTTATGGA GGACGTTTTT	4320
CAAATATTCT CCATCGAGGA CTAGGTTTAG GTCAGTGGAC GGATACAGCT GATGGTGGTC	4380
GTAGGCATAC TTTGTTATTA GACTATGCCA AAGGTAAAAA TAAAAAATGG TACGATCTGC	4440
AATTGCAACT GGACTTTATT TTTGACGGCG ATGCACCAGG TTCACGAACA GCAGCAGACA	4500
ATGTAGCAAG AAGTAAGGTG GCTGCAACTA TTCCAGAACT TACTACTTAT TATCTAACTG	4560
TTTGGGAAGG AAATCCCGBA GACAAGTTAG GAGAACGAAT TCAAGCTGCA CAAAATTGGT	4620
TTACATTCTT TTCTCGATCG GGTACACCAA TCGGTGGTTC TGGGAGTGAG GTCTTTGCGC	4680
AGTATAAAGA AAAAATGCAA CCATTACCCA CAGACAGAGA AmCAAAaGCA GGACaAGGCT	4740
GGCCAGGCMa TGCGTATGCG CCAGGGAATT GACTTGGTA TGTGTTTAAT CGTTTgCACA	4800
GTTAGGAAAA AGCATTTACC CTACTATGGG AAACGCCAAT CAATGGGTGC ATAACTACAG	4860
TCmAACGCCC GGAGCsACAC TTGAATCGGC CCCTAAAAAA GGGGATGCTG TTATTTTTAC	4920
AAATGGTGTA GCTGGTCTA GTACGCAATA TGGACACGTG GCTTATGTGG AGCATGTAAA	4980
CTCTGATGGT AGTTTTGTCA TTTCTGAAAT GAATGTCAGT GGTGAGTATT CTATGAATTG	5040
GCGTGTCTTA AAAAAAGAAG CAGGCGAATA CTTTATGCGC CTAACTAAA AAGAAAGGAA	5100
CCAATAAGAA CAATGTATAA TATGGAAAAA TTAAGTAGAT ACAAGAAAAT TGGCTTTATT	5160
GGTGGCAGCT TACTGCTTAT TGTGCTTATT TTTTGTATTG GGCTTAGTGT TGGGCAAAGA	5220
AAACAAGTAA ACACCAATGA AAAACAAGTA AAAGTCGAGA AAAAAGAGGA ATTAACAACA	5280
AGTACGGTAA AAAAATTTTT GATTGCGTAT TATACAAAAA AAGATCTTGG TGAAAATAGA	5340
AATCGCTATG AACCTTTGGT GACAAGCGCC ATGTATAACG AATTAGTGAA TGTAAGAAAA	5400
CAGCCTGTTA ATCAAGCGTA TAAAGGCTAT GTAGTCAATC AAGTACTTGA TACGTATAAA	5460
ATCTATATTG ATACTGAAAA CAATGAAGTC ATTGTGGACG TCACGTATAA GAATACACAA	5520
CGAACCAAAC GAAATAATGA TGAAGGAGCG TTGAAAAATC AGAGCAATCA AGAAGCGATG	5580
AAATTAACCT TTGTAAAACA gGCGCTAATT TTTTAGTGGA TAAAtGGCAC CAGTAACACT	5640

TACCAATGAA TTmCAAGAGG AGCCAAATTC TTATAATACa CATGTAGTAA CGACAGAAGA	5700
ATCAACAAAG GAGAGTGCAA ACAGTGGGGA AAAGTAAAGT AAGTGAATTA GAAAAAAGA	5760
AGGATCGAGT GGAGAAAGAG CTTAATAAAG CCAATGAACG TCTCGATACA TTGAAACAGT	5820
CGATTGCTAA AAAAGAGAAT GAATTGAAAC AAATAGAAGC GGAAATTGTC AGTGCTCTAT	5880
TAACAGAAAA TAATTTATCT TTTAGTGAAT TGA CTGCGTT GTTGAGTGAA AAAGAAaTCAG	5940
GTGGTGAAGC mAGTGTtTAT AGGGAaTAAa ACAACCTATC GCTATGAACT TCTTCATTCA	6000
AAAAATATTA AAGGCGAGGC GATCGACTCT ACCGCTACAC TTGTTCGTTA TGTGATTGGC	6060
GTAaAmCAaCG ATCTTATTAA GAAGGAAGAA GAAGGATGGC TTTATGTCGA AAAAAATAAA	6120
GAAGGAAAAC AaGAGGACCA AATCCTTTAT GGCAAGCGTA TAGATTTACC AATTTTAGAA	6180
GACAATCCTT ATTTTCATGA TCTTTTAGAT CCGTTTGATA CAAAAAAGT CATTCCTTTT	6240
GTAGAGCTAT TGCCTTCAGC TGTAAGAAAT CAGGTAAAGC CAGAAGAAAA AATTAAAGAA	6300
CCACAGGATA TTGTAGAGTT AGAGAAAAAT AACAAAACCA CTATTACTAC TGAAGAAAAG	6360
GAAATCACTG AATTAAAGA GCAACCAGAT GAAATGGCCG ATAAAGAAAC ACAGAAGAAA	6420
ATTCTGGAAC AAAAGATTCA GTTTAAAAA GACGAGTTAG CCACATTGAC CGAGGAGATT	6480
GCGACTAAGC AGGAAGAAAT CAATGACTTA GAACAAAAAT TAAGTACATT TTCAAAGAAG	6540
CCAGAGGGAC TTCATTTATA AAAAAGAGTT TTTAATCAGC TAGTAAAAA GCyTCTTCTA	6600
GCTAAGAAy AAGTATTATC ACTTmCCCT TACAGTTGTT TATGGCCTAT CAaAAAAGTA	6660
GAGAAATAAA AAAGACTaGC AAAAGTTGCT AGTATTTTTT TTGTAAAGGA GGAGAATTTT	6720
GGTGAACAA AGAGAAATTG TACACCGCTA TTATGTTTTA GGTACTAGTA CTTTAGATAA	6780
AGTATTAAAA GTGCTGCTTA ATTTATCCAG TGATGGTGTA TTGGCGCTAA AAGACAAGTT	6840
TCTTCCTTTA AAAGGGGAAA ATAATTTGTA TAAGCTAATG AATCGTGAGG ATCCACTGAC	6900
TTCGGCGCAA CTTCATGAAA AAGTGAATTT AAATAAGTTA AAAGAACAGT TGGAAGCCCA	6960
AGGGCTTCCT TTTGCGTTTA AAGAAACCAA AGAAGGGACT AACTTCTATT TTCGAGTAAA	7020
AGATACGGAA CTCGCAAGA AAGCCTTAGA ACGCGTATTA ACGGATATTA AGAAGAACCC	7080
ACAAATGATT TTGCGTAACC AAATACCATG ACTTTTGATG AACGGTTAGC CTACACTGCA	7140
GCCAATAAAA AATACGTGGG tAAAAtCGaT CAAACCAAAA GCcAtCcCcA AAGGGAAGGA	7200
AACTGTTGAT TGCnG	7215

## (2) INFORMATION FOR SEQ ID NO: 103:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5504 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

TGTTGCGCTA AGAGATTCAC CATAACCACT GATTGAAGTA ATTGGGGTAA CCCAATGCCC	60
CAATTGCTAA TCCCTCGAAT ATGCGCATCA AATTGACTAA ATGAGCAAGC TTCAATAGAG	120
TAATGACCTG AGTTATGAGG CCGTGGCGCT AATTCATTGA CGTACAACCTC TTCTGAATCC	180
GTTAAAAACA TTTCGATAGC TAAAACCCCA GATAGATTTA ATTCCTCAGC AATGTATTGC	240
GCCAATTCCG CTGCTTTTTG TTGCACATTC TCAGAAATTC GTGCTGGCAC AATTGATTCA	300
TGCAAGATAT TCTTCGATG AATGTTTTCA GATACAGGAA AGCTTTGGTA ACCATGCTTA	360
TTCCCTGCAA CGATGATTGA AAGTTCTTTA GTGAAAGGCA CCCACTTTTC TAATTCACAA	420
GTGGCGTTAG cTAATAATTT AGTCGCTTGG GAGAAATCCT CTTGCGATTT TAAGACGACT	480
TGTCCTTTTC CATCGTATCC ACCTTGAATA GTTTTTAAGA CACATGGAAA ACCAATCTTT	540
TCTATTGCTT GTGGCAGCTC TTTCTGGCTT TTTACTTCCG CAAAAGGGGC AATCTTACAG	600
CCAATCGATT GAAGAAACGT TTTTCAAGC AATCGATTTT GAGTAATTTT TAACAATTCT	660
GTTCTTGTTG GCAACTTTTC TAGTGGAATG GCCTTTTCTA GGGAGGCCAC TTCAATATTT	720
TCAAATTCAT AGGTGATCAC ATCGCTTAAC TCTGCTAGTT CAGTTAAAGC AGATATCTGA	780
TCAAAGGCTG CCACAATTTG ACGGTCAGCT ACTTGACCAG CTGGACTATT GAGCGTTGGG	840
TCCAACACAA TCATTTGGA ACCCATTTTT TTACCACTCA TGA TAGCAT TCGCCCTAAC	900
TGTCCCCCAC CTATAATGCC AATAGTTGCT CCTACTTTTA AAGGTTTATT CAAGTTGTTT	960
CCCACTTTCC ATAACCGCTT TTTCCAATGA TTGTCGTCTT GCTTGGATTC TTTCTGCTAA	1020
CTGTAAATCA TCAATACTTA GTATTTCTGC AGCTAATAAC CCAGCATTTG TAGCACCTGC	1080
TGAGCCAATC GCAGTTGTTG CTACAGGAAC TCCCCCTGGC ATCTGCACAA TAGACAACAG	1140
TGAATCAATA CCGTTTAACG TTCGTGTTTT AACTGGAACA CCAATGACAG GTAAGGTTGT	1200
CTTTGCAGCT ACCATCCCTG GTAAATGGGC CGCTCCACCA GCGCCAGcgA nTAATAACCT	1260
TCAATCCTTT TGTTCTGTGAT TCCTCTGCAA AACGAAACAT CAAGTCAGGG GTCCTATGTG	1320
CTGATATCAC TCGTTTTTCA TAAGGAATTT TCAACTCTTC CAACACAAGA CAAGCAGCTT	1380
TCATCGTTGC CCaGTCGTGAG GTACTTCCCa TAATCACAGC GACTAATGGT TtCATGAATG	1440
TAATAACTCC TTCCGCCTTT TTTCACTGTT ATTCTAACAA TGAGAAAAAG AGATGTCAAr	1500
GAAAAaCCGA AATATTATAG ATTAATAATTA TATAATGTTT GTGTTTATCA AAAAAAGAGT	1560
CATCGGAAAG ATgcTTTTTT GctATTTTTA GCGTTGCTTC CTCTTTAAGA AGAAACGTCG	1620
CCTGTATTAA TAACTGGTTG TGTGCCCTTA kCAGTAATGA TTGAGACTAG TAAATTATTA	1680
ATAAGTTGTA CTTTTCGTTC ATCGGTAAAA TTAATTTCTT GCCCTTCTTC AATCTGTTCT	1740
AAGGCCATTT GTGTCATGCT TACCGCACCT TCAACAATTG TTTGTCGAGC CGCTAAGATT	1800
GCTTTTGCTT GCTGTCTTTG CAACATGGAG CTAGCTATTT CAGTTGCATA AGCTAAATGA	1860
TTCaACCGTG TTTGATAAC CTCTACACCA GCCACTGCTA AGCGTTCTTG CAATTCTTTG	1920
GTAACTCCT CGGAAATTTG TTCGGTATTT CCTCGCAGAG TTACATCATT ATCTGAAAAA	1980

GTGTCATAAG GATATTGTGT TGCTACATGC CGGATTGCCG TTTCACTTTG AATTTCCACA	2040
AAGTCTTGaT AATAATCGAC ATTAATAAAC GCCTTAGCTG tATCGACyAC CCGAAAAACG	2100
ATGACCGCAG AGATTTTCGAT TGGATTCCCG TCAGAGTCAT TGACTTTTAA TAGAGAACTA	2160
TTGAAATTTT GCACTTTCAA AGAAATGTTT ATTTTTTGCG TAAATGGAAT GGTAATAAAC	2220
AACCCATTTT CTTTAATCGT ACCTAAGTAG CGACCAAAAA ATAAAAATAG TTTCGCCTGG	2280
TTGGGACTAA CAATGGTTAA GGAACCTAAA AATAAAATAG CCCCAACTAA CAAAATAATA	2340
CCTAAAACAA CAAGAACACC ATTGGTATGA CTCGCTCCAT AAATAAATAA ACCAATTCCC	2400
GCAATTAAAA TAACTAATAA GCTCGCAATT CCTACGTAGC CACTCATATA AAATGTTTTT	2460
TTCTCTTCCA TCTGTTTCAC TCTCCTAACT CATTATCTG GTCCTTATTA TACACCTCAT	2520
TGCCTTCAAT AGAGTAACTT TTAACATTAA ATAAACTGAT TTTTTCGGGG CTCTTGAGT	2580
ATACTGAGAA CAAAGAAATC CCAAGAAAAG GAGTGATAGA ATGCGCTATT TAAGCAAGGA	2640
CATTTTAGAG GAAGTCATCA CTCAACGCCC CTCTGACTCT TATAAAAGTA ATTTTGGGCG	2700
TGTTGTATTA ATCGGAGGAA ACCGACAATA TGGCGGAGCa TCATCATGAG TACCGAAGCG	2760
TGTATCAATA GTGGCGCTGG TCTCACCCT GTGATTACTG ATGTTAAAAA TCACGGGCCT	2820
TTACATGCAA GATGCCCAGA AGCGATGGTC GTGGGCTTTG AAGAAACAGT CCTTCTGACG	2880
AACGTCGTAG AGCAAGCGGA TGTATTTTA ATTGGTCCAG GTTTAGGCTT AGATGCTACT	2940
GCACAACAAA TATTAAAGAT GGTACTCGCC CAACATCAAA AACAACAATG GTTAATTATC	3000
GATGGCTCAG CAATTACCCT GTTAGCCAA GGCAATTTTT CCCTTACCTA TCCTGAAAAA	3060
GTTGTTTTCA CACCTCATCA AATGGAATGG CAACGTTTGA GTCACCTACC TATCGAGCAA	3120
CAAACACTGG CAAATAACCA GCGCCAACAA GCAAAATTAG GCAGTACGAT TGTTTTAAAA	3180
AGTCATCGCA CCACAATTTT CCACGCAGGA GAACCTTTTC AAAATACAGG TGGCAACCCT	3240
GGGATGGCTA CTGGCGGAAC TGGGGATACC TTGGCTGGCA TCATTGCTGG TTTTTTAGCT	3300
CAGTTTAAGC CGACAATTGA AACCATTGCT GCGCGCGTCT ACTTACACAG TCTCATTGGA	3360
GATGATTGG CCAAACTGA CTATGTTGTT TTACCAACGA AGATTAGTCA AGCCTTGCCG	3420
ACATATATGA AAAAATATGC CCAGCCCCAC ACTGCACCTG ATTCTGAACT ATTAGAACAG	3480
AAGCGTTTCA GATAAAACCC GAAAAGGCTA AATGGTGACA AATTTTGGGT TAAAAAGCAG	3540
GATTTAAGCC TGGGCCGAAA ATCAAAAGCG ATTTTCAGCC CAGGCTCTCT TTTTATTTT	3600
CCTTATAAAC ATTTTTTAAC GTTAATGGAG CGCCAAAGGT ATGAAAATC AAGCCTTTGA	3660
CTTTATCATT TTGCAGGACT GTTGTTCAT TTtGGCTAAG TGGAATCATG CCAGCAGTCG	3720
TTTCAATCAC TTCTTTTTCA GCTGCAATCA GTGTCGCCCA TCTTTTTTCT GGCTCTAAGG	3780
CATAGGTTGT GGCTGCTTCA TCTAATAATT TGTCATAGAC AGGGTTCTGA TAATTGCGAT	3840
CATTGCCCTT GTATAAAGTC ATCAGGGTAG AAATAGGGTC TTGATAGTCT GGTGTCCAGT	3900
AAATTAAGAA TAAATCATAG TCACTTTCTC GCCCAAAGTT CAATGCAGCT TCAGTCGGCA	3960

ATGCGGTTAG CTCTATTGTC AAACCAGGAA ACAATTCTTG TAGCGAGCCT TGCAAACTTT	4020
CACCAATTTT TTTGTAAGAA CCATCATCTG TTACCATCAA TTCAACGTTA ACCTTTTCTC	4080
CTAATTCTGC TTGTGCTTTT TTCCAATAAC TTTGCGCTTT TtCTTTGTTA TAAACCATTA	4140
AATTACCTGC TTCTTGACGA AAATCGAGAC CCGTTGTGGG ATTCGCCACA AAGCCTTCCG	4200
TAATCGCACC ATGTAGCGCT TTAGAACCAT CTGCAATAAT ATTATtGACT AGATTTTCTT	4260
TATCTATTCC TAAAGCCAAT GCTTTACGCA GGTTTTCGTT TGCAAGCGGC GTkGCTTGCC	4320
CTTyCCGTTT TTGATTAAAG CGCAAATAGT TCATTGTGCG TGTTGGATAC GAATGATACA	4380
ACGTATTATT TTTATTCTGT TCGGCCAGTT CACCACCTAG TGTAGCCACA TCTAATTGTC	4440
CATCTTCAAA AAGATTCAAT GCGGTAGATG TTTCTTTGAT AACTGTATAA TTGATAATGT	4500
CTGAGCGCAC GTTCTGGTGA TCCCAATAGC GATTATTTTT TGCTAGTTGC CAGTCCATCT	4560
TTGTTTGCTG CCAATCTTTT ACCACGAACG GACCATTATA GACGACTTTA TCACtAGCAG	4620
TTCCATAGTC CGCACCAAAT TGTTGACTA CTTTTkGATT TtGCGGGaAA AATGTCGGAA	4680
AaGCTAACAA GGACGTAAAG TACGGTTTTG GCTCTTTGAG CGTCACCTTT AATGTATAAT	4740
CATCCACAGC TGTGACACCT AGTTCATTGG GTGCTAATTT CCCCGCTGAG ATTTCTGCAC	4800
CATTTTGAAT TGTTCAACG ATGAGGAAGC TATAAACAAA GCCGTTTTTA GGATCAATCA	4860
TTTTTTTCCA AGCATATTCA AAATCATGTG CTGTGACAGG ATCATCGTTA CTCCAAACCG	4920
CTTCTTTTCT CAAAGAAATC GTGTAGGtTT TTCCATCTTC TGaAATCATC GGCAATGCTT	4980
TTGCTACGGC TGGTACCAAT TGGTCTTGTT CATCTAACT ATATAACCCT TCAAACGCTG	5040
CAGTTTGAC AATAGCATCT GGAAAATCCA ATAATACAGA GGTGTTGAGC GTTGTTAATT	5100
CCGAGGGTGA CATCAGTTGA ACGGCTTCTT TGGTATTTTT CTGTTTCTTT TTATTTTCAT	5160
TTTTATTGGT ACACCCAGCT AAAAATCCTA CCAATAAAAT ACAAACCATT GTGATTTTTA	5220
AATACTTTTT CATCTTCTTT CTCCTATAAC TCTTCATTTT TTTATGCTCT TACTTACAGA	5280
AAGACAGCTC CCCTTTTATT TTGTCCTTC CACAAGCATC CAGACAACGA CCGCTGGCTC	5340
CAACCCATTA TTATAATTAA TGTGCCCTC ACCAGCTTCA ATTAATTGGG CTTGTCCAGC	5400
TTTCCAACCT AAGTCTTGAC CATCTTCTAA AATCGTATGT GTcGCCAGAA ATCACATAAG	5460
AATATTATC CTGTTTCATGG CGAGCATATC CAGCTAAAGG CCGT	5504

## (2) INFORMATION FOR SEQ ID NO: 104:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6433 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

GTTGCGACTA AACTATAGCC AATTGCTGAA AAACGCTTAG CAATTTCTAA CGCCTCTTCT 60



TTTGTTCAT CAGCAATGGT GAAAAGTACA GCCCCATAAC TTGGCAAATG CAAGCCAGAT	120
GCTTCAAAGG CTTTATACAA GGCTTTTTCT AAATAATAAT CACTACCCAT AACTTCCCCA	180
GTTGATTTC TTTAGGCCC TAAGTACGTA TCTACTTTTT GTAGTTTCGT AAAGGAGAAT	240
ACAGGGGCTT TCACATGAAC TTGTTTACTT TCAGGATATA ATCCATCTTG ATATCCTAGG	300
TCAGTCAGTT TCTCACCAAG AATCGCTTTT GTAGCAACTT GGGCCATCGG AATCCCTGTG	360
ATTTTACTTA AGAAAGGAAC TGTTGACTC GCACGAGGAT TCACTTCAAT CACATAAACA	420
CGATTTTCAT GAATAACAAA TTGAATATTC ATCATACCTA CACAGTTTAA ACCAAGTGCC	480
AGTTTTTTTG TATAATCGGC AATTGTTGCT TGAATTTCTT GAGAAAGATA CTGCGGTGGA	540
TAAACCGCCA TTGAATCACC AGAATGCACG CCAGCTCGTT CAATATGTTT CATAATCCCT	600
GGAATTAGTA CTGTTTCACC ATCACAAATG GCATCAACTT CACATTCTTG TCCCAATAAA	660
TAGCTATCAA CCAAGACAGG ATGTTCTGGA GAAGCTTTGA CTGCATGTCG CATATAATCT	720
TCCAAATCCT TTTGATTTTC GACAATTTCC ATTGCACGAC CACCTAAAAC ATAAC TTGGA	780
CGAACTAAAA CGGGATAGCC GATACGATCA GCAATCACGA CTGCTTCTTC TGCAC TTGTG	840
GCCGTATCTC CTGGTGGTTG AGGAATCGCT AACTCTTGCA AAGCTTGTTT AAATAAATCA	900
CGATTTTCCG CACGATCTAA ATCTTCAATC GTTGTTCCAA GAATTTTAAC GCCTTGTTTT	960
GTGAGCGGCT CAGCTAAATT AATCGCTGTT TGTCCACCAA ATTGGACGAT CACGCCAATC	1020
GGATTTTCTA AATCAATCAC ATTCATGACG TCTTCTAAAG TTAATGGCTC AAAGTATAAT	1080
TTATCAGAAA CAGAGAAATC TGTTGAAACT GTTTCAGGAT TACTATT CAT AATAATTGCT	1140
TCATAACCAG CCGCTTGGAT CGCTTTGACA GAATGAACCG TTGCATAATC AACTCGACC	1200
CCTTGACCAA TGCGAATGGG ACCAGAACCT AATACTAGTA CAGAAGGCTT TTTCGAAACA	1260
TGCTTTTCGT TTTCCACTTC ATACGTGCTA TAAAAATAAG GCGTGTGTGA TTCAAACTCT	1320
GCAGCACACG TATCCACCAT TTTATAGACG GGCACAATTT GATTGGCTCT GCGGAAGTCA	1380
GCGATGGCTT GTTCTGTTTG TCCCCAAAGT GCTGCGATTT TCCGATCAGA AAAACCATT	1440
TGTTTGCTT CTTTCAAAAC AGCTACGTTG TCCACGTGAC TTTCAAGGGC TGTTTCAATT	1500
TCGATAATGT GTAAAAGTTT ATCCAAGAAG AATAAATCAA TTTTCGTTAA ACTTTGTAAT	1560
TCTTCGATGC TATATCCTCT GCGAATGGCT TCTGATAAGT AGAACAAATCG ATCATCTTGC	1620
GCATGGACCA TTTTCTTCGT TAATTCCAAA TCACTGACAT GACTTAGTTC TGCTAACTCG	1680
TTGTGATAAG CACCAATTTT TAAAGAACGA ACTGCTTTTA ATAAGGATTC TTCAATATTG	1740
CGGCCAATTG CCATGACTTC CCCGGTTGCT TTCATTTGCG TTCCTAGTTC TCGAGCACCT	1800
TTTTCAAATT TGTCAAAAGG CCATCGAGGA ATTTTGAAA CAACATAGTC CAAAGCTGGT	1860
TCAAATTCGG CATAAGTGGT TCCAGTAACT GGATTTTCA TTTTATCTAA TGTTAAACCA	1920
ACAGCAATTT TTGCAGCTAA CTTGCAATG GGATACCCTG TGGCTTTACT AGCAAGTGCA	1980
GATGAGCGAG AAACACGAGG ATTCACTTCA ATcACaTAAT AwTTAAAGcT ATGCGGGtCC	2040

AArGyTAATk GAACwTTACa GCCACCTTCA ATTTTAAATG CACGAATaAT TTTCAArGAC	2100
GCATCTCGCA ACATTTGGTA TTCATAATCT GATAAGGTTT GACTCGGAGC AAAACAATC	2160
GAATCCCCAG TGTGAATGCC GACAGGATCA AAGTTTTCCA GTTACAAAC CACAATCGCA	2220
TTATCAGCCG AATCACGCAT CACTTCATAT TCAATTTCTT TAAAACCAGC AATACTTTTT	2280
TCAATTAAAC ATTGTGTAAC TGGTGATAAT TTCAGACCGT TTTCAGCAAT TTGGCGGAGT	2340
TCTTCTTCG TATCACACAT GCCGCCGCcT GTTCCACCTA AAGTAAATGC GGGACGAACG	2400
ATAATTGGAT AGCCAATTCG TTTGGCAAAA GCCACGGCTT GTTCCACCGT ATTGACAATT	2460
TCACTTTCTG GAATCGGTTG TTCTAATTCT TCCATCAATT GTTTAAATAA ATCACGGTCT	2520
TCCGCTTGGT CAATCGCACT GAGCTTTGTT CCTAGCAATT CTACATTTAG CTCATCTAAA	2580
ATTCCTGATT CCGATAATTC CATCGCCATA TTTAATCCAG TTGACCACC TAATGTTGGT	2640
AGAAGAGCAT CAGGACGTTT TTTTCGCAA ATTCTTGAAA CAAATTCTAA AGTAATCGGT	2700
TCGATATAAA CATGATCCGC GATTTCTTTA TCGGTCATGA TAGTCGCTGG ATTTGAGTTT	2760
ACTAAGACAA CTTCATACCC TTCTTCTTTT AGCGCTAAAC AGGCTTGTGT GCCTGCATAG	2820
TCAAATTCTG CTGCTTGTCC AATAATGATT GGTCTGAAC CAATCACCAT GATTTTCTTG	2880
ATGTCCGTTT GTTTTGGCAT TAGTTTTGCT CCTTCCATGC ATCCATTAAT TCCATAAATT	2940
CGTCAAATAA ATGAAGACCA TCGTGCGGTC CTGGAGCCGC ATCTGGGTGA TATTGCACAG	3000
TAAAAGCAGG ATAATCTCGA TGACGCACGC CTTCAACCGT TCCATCATTG ACTTCCACGT	3060
GGGTACAAG TAATTTTTCT GGATCAATTG TTGCTTCATC TACGGCATAT CCGTGGTTTT	3120
GTGAAGTGAA ATCAATGCGT CCTGTTGCAA TTTCTCGTAC AGGATGATTC AAACCGCGAT	3180
GTCCAAACTT CATCTTATAC GTATCCGCAC CATTGCTAA TGAAAATAAC TGATGACCTA	3240
AGCAGATGCC GAAAATCGGA ACTTTTCCTT GGATTGCTTG AATCATTTCA ATAGCTTCAG	3300
GCACATCTTT GGGATCTCCA GGCCCATTTG TCAACATGAC ACCATCGGGA CTTAATTCTA	3360
AAATCTCTTC GGCCGTTGTA TTATAAGGCA ACACGGTTAA GTTGCATTGC CGTTTTGATA	3420
ATTCTCGCAA AATACTATGT TTCAACCCAA AATCAACGAC GACCACATTA CGCCCAATAC	3480
CTGGACTCGG ATAAGGCTTA GTGGTGGAGA CTTGTGCCAC TTGATTTTTC GGCATGACTG	3540
TCGCTTTTAA TTGATCAAAC GCATGCGGCA AATCATCCAC GGCATCAATG ATGCTCCCTT	3600
TCAkTGTTCC AGCTGATCGT AACTTACGAG TCAGCGCACG CGTATCAATT CCTGAAATTC	3660
CCGGAATCCC TTTGCGCTTT AAAAATTCAT CCAAGGTCAT TTGTTGACGC CAATTTGAAG	3720
CAACCCGCGC ATGCTCTTTA ACAATGACAC CTTTACAGGT TGGTGCaATG GATTCATaAT	3780
CaTCTCGGTT GACCCCGTAA TTCCCaACTA ATGGATACGT AAAAGTAATC ATTTGTCCAT	3840
TAAAACTTTG GTCAGTAATT GCTTCTTGGT AACCTGTCAT GCCAGTAGTA AAAACAACCT	3900
CTCCTACTAC ATTGCCtTCT GCaCCAAACG CTTTTCCTTC AAATACTGTT CCGTCTTCTA	3960
AAATCAATAA CCTTTTCAA CCTACTGCGC CTCCTCCGAC CAAGCTAATG CTCCGTCAAC	4020

AAATGTCATA	AGTGTCTTTC	CTTTGACTGT	CCACCCTGTA	AATGGTGTAT	TGACAGCCAT	4080
TGATTCAAAT	GCTTCTGCAT	CAATTGGTGC	AGCTGTTGCT	AAGTCAAAAA	CTGCAATATC	4140
TGCTGGTGCA	CCAATTsTTA	GTGTTCTGCG	GTTTAATCCA	AAAATTTCTG	CCGGTTTCAC	4200
TGCCATCCAA	TCAATCACTT	GTTCTAATGT	GAAAATGCCT	GTTTCGACAA	AGTTCGTATA	4260
AATCAATTGG	AAGGCCGTTT	CACTACCGAC	AATACCAAAA	GGTGCCTTTA	AAAATGATTG	4320
TTGTTTTTCT	TCTAAACCAT	GGGGCGCATG	GTCGGTAGCA	ATACAGTCAA	TCGTGCCATC	4380
TAACAAGCCG	TCAATCAATG	CTTGTCGATC	TGCTAAACCA	CGTAATGGCG	GGTTCATTTT	4440
CCAAAACCCT	TCGTCAACCAG	GAATATCCTC	ATCGACTAAA	ATTAAATGGT	GTGGCGAAAC	4500
TTCGGCTGTT	ACGTGAATCC	CTGCTTTTTT	AGCATCTCGA	ATAACGCGCA	CGCTTCTTTC	4560
TGTTGAAACA	TGACAAACAT	GATAATGAAC	GCCTGTTTCT	TTGCTAAAG	TAATATCCCG	4620
AGCAATTTGT	GAAGCTTcCG	TTGCACTTAA	AATCCCTGGT	AGGCCTAATT	TCTTAGAAAC	4680
TTCTCCTTCA	TGCATCACAC	CACCAAATAA	TAAAGATTCA	TCTTCCGTGT	GCGCAACAAG	4740
TGCCATATTT	AATGCAGCCG	CTTCTTTCAT	GGCTAAGTAC	ATCGTTCCTG	CCGTTTGAC	4800
ACCAACACCA	TCATTCGTAA	ATGCAAAGGC	GCCCCTTCT	TTTAAGGCCT	TCTGATTGGT	4860
TAAAACTTCA	CTCCGTAGCT	CTTCTGTAAT	GGGGGCATAT	TGTAAAACTT	TCACGACTGC	4920
ATCTTTTTGA	ATTAAGTCGT	AAACCTCACT	TAATTTTTCA	GCCGTATCTG	GTACTGGATT	4980
CAAATTGGGC	ATTGCGCAAA	CCGTAGTAAA	ACCACCGCGA	GCAGCAGCTT	TACTACCCGT	5040
TTTAATCGTT	TCTTTATACG	TAAAGCCTGG	TTCTCTAAAG	TGAACATGCA	CATCCACCAA	5100
GCCTGGCGTG	aTTAGTTGCC	CGTTCGCATC	AAACACTTGC	TCAAAAtCTG	CTTCGTCAAA	5160
CGATTCCCCG	ATTGCGTGAA	TCACACCATT	TTCTAGCCAA	AGTGCTGCCT	CAATCAATTG	5220
ATTCTCTTTT	TTGATGATTT	TTCCATTCTT	AATTAACGTC	TTCATTTCTC	CACCTAGGCC	5280
TTTCCATGTA	AAATTGCTTC	CAAAATTGCC	ATGCGCATAA	AGACTCCGTT	ACTCATTGTC	5340
GCGACGATTC	TTGATTGCAG	ACTTTCCTCT	AATTCATCTG	CCAGTTCAAC	ATCCCGATTG	5400
ACTGGCGCTG	GATGCATAAT	AATTGCGTGT	TTCTGTAAAC	GAGTTGCCCG	TTCGTTGGTT	5460
AAACCATATT	CCAAATGATA	GCCTTCTTTT	GAAAACTTT	CCTTACCATC	ATGACGTTGC	5520
TGCTGCACGC	GAACaACATC	ATCACATCGA	CTTTTTC AAC	AATTTATCT	AGCGGCACAT	5580
ATTGTCCATA	GACATCAAAC	TGATGATCGT	ACCACTCTTC	TGGACCTGAA	AAATAAATTT	5640
CAGCGCCTAA	ACGGTTTAAT	AATGTCATAT	TTGATTTAGC	AACGCGAGAA	TGTGTAATAT	5700
CTCCCACAAT	CGCCACTTTC	AAGCCTTCAA	ATCCACCAAA	TTCTTCATAA	ATAGTCATCA	5760
GATCCAATAA	ACATTGTGTA	GGGTGTTGTC	CACTCCCGTC	CCCACCGTTA	ATAATAGAAC	5820
ATTGAATCGT	TTTACTTTGA	ATCAATTCAT	CGTAATAATT	TTCTTTACCA	TGACGAATGA	5880
CTGCTACATC	TACACCAATT	GCCGACATCG	TTAGTACTGT	GTCGTATAAC	GTTTCTCCTT	5940
TTTGAACAGA	ACTTCTACTT	GCTTCAAAC	CAATCACTTC	TAAGCCTAAT	TTTTTCTCAG	6000

CTACCTCAAA ACTTTTATGT GTTCGAGTAC TATTTTCAAA AAACAAGTTT GTTGCAAAAT	6060
ATTGACGTTT TTCTGGGTGC CATTTTGCCC CCTGTTTAAA TTCTCCTGCG CGACGGATTA	6120
ACCCCATGAC TTCACGGTCG GTTAAAGCCT CAGCAGTTAA AAGATGTTTT AAAC TAATTC	6180
GTTCAGATGT AATAATCATT CGTCCAGTCC TCCTACTGTT CTTCGCTACG TGCCTTTTCC	6240
GGTAAATTA AATTCAAAAC AATTCCTAAG ACAGTTGCTA ATGCCATTGA AGACAATTCA	6300
AAAGTTCCCTA CTTTGAAAGAC TAAGCCACCA ATGCCGATGA CTAAAATTAC TGAAGCAATC	6360
AATAAATTCT TTTTCTTATC AAAATTAATC TTATTATCGA TTAAGATTTT CAACCACTGG	6420
GGnAACAAAC CGA	6433

## (2) INFORMATION FOR SEQ ID NO: 105:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5855 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

ATAGAAACCG CCTCTCTCTC GACACGCAAC TTGCGTGTTA TCTGTACGAT TTGTTTACAT	60
GTTGCATTGT AAACAGTTTT TAGGCAATCG TCAAGCAGAA AATAATTAAC gTTGCAtGAA	120
GAAACtAATA ATTGCAAAAA TAGCAATGTG AGCTGGATAG AAAATATAAA AGAAATTCGG	180
CATGCCACGG CTTTTTTCAC CATTATAAAG AACAATTGGA ATCGCTGCTA AAATCATCAT	240
CCATTGGTGG TTCACCGGGA ATAGTTCTTG GAAATTATAG CTGGCTGCTG AAATCAACGC	300
AACAACGCCA ATTGCTGAAA TTTGTAACCA CTTTTTGCCA TGGAATAAAT AAAAAGCTAC	360
GGCCAAGAGA ACAAATAAAA ACCCACCTTC TGTCATCATT AGTGATGGCA CAATCAGATT	420
GAAAATTTGG AAACCAATTA AGATAGCAGG GTTGCTTAAT ATGTCTGGAG AACTAAAAAC	480
AACGAGAGGT AGGGCACTGA TTAATAAAGG CACAATAAAC CAAAGAAGCC CTTTACCAAT	540
TTGTTTTTGT TTAAATAGT CCATAGACTG CATGTAGACC GTACCTAAAA ATAACGTCCC	600
AAAAATATTA TTGATAATTA AGCCACCTGT TGAAAAATAC GCATTTAAAA AGCCGTTTCAG	660
AATCCCCATA ATCCAGAAGC CTAACAATAA CCGGAACATG TATTTTCGCC GATTGCTTGT	720
GTGAATGAAG CCTTCAGAAC TTTCAAATAA GAAAATAGGT GCAGCAATTC GTCCAATCCA	780
TCCGAACCAA CCAGGCACGC CTAAAAAACT AAAAAATTGC AATAGATGGT CAAAGACCAT	840
TGAAATAACC CCAATGACTT TCAAATGAAA GCCAGTTAAT GTTTTACCCA TAGTATAAAC	900
GCCTCTCTTT CATAAAGAAC AATATAACAA AGTCCAGAAA AACCAACCAT TTATTTGCCT	960
AACAGTTATC TTACAAAATT GTTTGATTTC TTCAGCTTTT GATGAAAGAA AAAAGTGCCA	1020
AGACTAGTTT TTATAATTAT TCTATGCTAT AGTAAAAGTA CATTTTAATA AAAAACATTT	1080
GGGGTGCTGT TATGGCTGAG ATGATACCCA TTGAACCTGA TGCAGTTAGT ACTGTCGCAG	1140
GGAAATGCCG ATTTACTTTG ATTTTTCACG CACTCATCTT GTCTTAGGAT GAGTGTGTTT	1200

TCTTTTATCA	GGGTTTATTT	AAGTCACTCC	TTTTGCTTTT	TGTTAAGAGA	AAGGAGTTTT	1260
TTTATTGAAA	AAATGGTCAT	TGCAACAAGT	CGTCTTACTC	GCTTTTTTTAG	CCTTTTTTGT	1320
TGGCGGTGTT	TTTATGGGGG	CAGGATTCCT	CTATGCGTTA	TTGAATGCTG	CGCTGTTACC	1380
ATTAGGACTG	AGTCCATTTG	CCAATGAGCT	TTTATTTGGA	ATGTGGACCA	TGGTGGCACC	1440
AATTGCAGCG	ATGCTGATTC	CACGAGCGGG	AAGTGCTGTT	TTGGCGGAAG	TTTTAGCCGC	1500
TTTGGCAGAA	ATGCTTTATG	GTTTCGTATTT	CGGTCCTAGT	GTATTAATTT	CTGGTGTGAT	1560
TCAAGGTCTA	GGTAGTGAAA	GTGGCTTTTT	GGTGACACGC	TACAAACGGT	ACGATACACT	1620
AACGTTATTT	TATAGTGCGA	TTGGGACGAC	TATTTTTTAGT	TATGTGTATG	AATACTTTAA	1680
GTTTGGTTAT	GGCAATTACG	GTCTGGGGAT	GAACATTGCT	TTAATTAGCG	TTCGGTTCGT	1740
TTCTATTTGT	TTCTTTGGTA	TTTTCTTAAC	GAAAGTTATF	TTACGAATGT	ATCAGTCTGC	1800
GCAAGGGTTA	GCTGTGAAGG	CAAAATGAAT	AGTATCCAAT	TAAAAGCTGT	GTCCTTTTCA	1860
CGAGAACAAC	CCTTGTTTGA	AAAGACAAAT	TTAGCAATTC	CTCAAGGAAC	GTTTTCTTAA	1920
TTAATTGGCG	ATAGTGGTAG	TGGAATAATCA	ACGCTCTTAC	GATTGATTGC	CGGATTTGCT	1980
CCGCTTGATT	ACCAAGGAGA	AATTTTGATT	GAAGGAATGG	AACGTCGACA	ATTATCAACT	2040
CGTGAAAAGG	CGCAAAAAAT	TGGCATGCTG	TTTCAAAACC	CCAGTCAACA	ATTTACAATG	2100
AAAACCTTGG	AACGAGAACT	GATTTTTGCT	TTAGAAAATT	TAGGAATCCC	GCCTGAAGAA	2160
ATGAACCGGA	AAATCCAAAC	AGCGCTTCAA	CTTGTTTCTA	CTCAAACCTT	ATTCACACGA	2220
GAATTAGCGA	CGTTATCTGG	TGGTGAAAAA	CAAAAAGCGG	CGTTGACAGT	TCTGCTAGCA	2280
ATGAACCCAG	ATATTCTTTT	ATTAGATGAA	CCTTTTGCAA	GTATTGATCC	AACCTCTCGA	2340
AAACAATTTA	TTCAAATATT	AGCAAGACTC	CATCAAGCAG	GCAAAACGAT	TTTGGTGTGC	2400
GACCACGATT	TCAGCGATTA	CGCTGATGTT	GTGGACCAAG	TCGTACAGTT	GAAAAATGGC	2460
CAGTTTGAAA	AGCAACCGTT	GACGTTTATC	AAAACAAAAC	CACAGACATT	TCAGCTTACA	2520
ACTTCTGTGG	TAAAACAACC	AATGCTCCAA	TTGAAAAATT	TTCGTTTGTC	TCAGGGAAAA	2580
CGAGTGCTAC	TTGAAGAAAA	AGAAGCGCTT	TTGTTTAAAG	GAATTACCAC	GTTGACGGGG	2640
CCAAATGGGG	CGGGAAAAATC	CACCTTATTA	AAAGCCATCG	TTCAAGAGACA	AAAATATCAA	2700
GGGAAAATGT	TCTTGGCAGG	CCGCCGCTTG	CGGGCTTCAA	AAAAATTGTA	TCAACACATG	2760
ACATTAGCGG	TGCAACAAGC	CAATCGTCAG	TTTGTCTACG	TGACTTTACG	AGAAGAGCTT	2820
TTGTTTGGTC	AGAATCATGAC	AGCAGAAAAA	AGAAGAAAAC	AAGAGGAAGC	CTTAACCTTT	2880
TTAGGTTTAA	AAGAAAAACT	AGAGCACAGT	GTTTTTCAAC	TAAGTGAAGG	GCAAAAAAAG	2940
ATGGTGCAAT	TAATTAGTTT	GCTGAGCTTA	GATTTGGACT	GTTTACTTTT	AGATGAACCT	3000
TTTGCTGGAT	TAGATGAGCG	AGCATGCAAT	TATTTTGTTG	AATGGATCAA	AGAAAAAAGT	3060
GCGCAACAAG	ATTTTCTAAT	CGTCACGCAT	CGGCTAGAAC	CTCTCTCTGG	CGTGAGCAAT	3120
TACCGAATAG	AATTGCTGCA	GCAACAACCTG	ATTATCTGGC	AGGAGGGAAC	CACATGCAAA	3180

TAAAGCAAAC	AAATGCAGCA	ATTTATGCTT	CTTTGATTCT	TATTTTGACC	TTTGAATTAT	3240
CTTTTCTCA	GTCAATTTTG	GCAAATTTGG	CGGTTTTTAT	GGGTTGTGTC	ATTTTTTTAA	3300
TAGGCCAGCG	AAAATCCCGC	TTACTGCTTT	GGTTATTTTT	CTTGCCACTT	TTACCAGCAA	3360
TCGGGACGTT	TTGGTCTATT	TATTTGCATG	GTACGAGTAG	TCAACAAGCC	TGGCTTTTGT	3420
TTAGCCGAAC	GTATGCCTTT	GCCGGTTTAG	GACTAGCTTT	TGCTGTAGGT	GTTGATTTTG	3480
AAGAATTATT	GTTGTTATTG	GAACAAAAGG	GGCTAGCACC	GAACTTTGTT	TATGGAATTT	3540
TAGTGGTGGT	TCACGCCTTA	CCGGAAGTGA	AAAGAGAAAT	TAATGATTTA	AAAGAAGCAA	3600
GCCTACTCCG	TGGCAAAACG	TTTCATTTTT	GGTCGCCGAT	GCTTTATGTT	AAAACACTGC	3660
TCGTGCTGT	TTCTTGCGCA	GATAAATATA	CTGAAGCAAT	GTCTGCCCAT	GGCTATGAAG	3720
AAGGAGCCAC	CAGAAGTCGA	AAAGAGCATT	TTGTGTcAGC	AAAAGTCAGT	CTGGGGCTTG	3780
CTATTTTGAT	CATCATTTTT	ACAAATTTGT	TTATTTTTCT	AACGTAAAGG	AGAAACCTAT	3840
GACATTTACA	GAACAAGCCA	AAGATCaAGC	GCAAGCTTCT	TGGCAAGGGA	GTTTTTCAGCA	3900
TCCTTTTATC	ACAGAATTAC	ATGAAGGAAC	CTTGAGTCCA	ACGATTTTTC	GCTACTATTT	3960
AATCAAGAT	CATTACTATT	TAAAACACTT	TAGTCAACTT	TATCGCTTGA	TTGCGGCACA	4020
AACGCAACAA	CCTCGATTGA	AAAAATTACT	CTTAACAAAT	GCGGAAAATC	TCGCGTTGGG	4080
CGAATTAGCT	ATTCGTGAAA	CTTTTTTTGA	AGAACTAGCG	ATTACGGAAG	AAGAAGTTGC	4140
AGCGACACCG	ATTGCACCAA	CCGCGTATCA	CTATGTTTCT	CATATGTATC	GCCAACTTAT	4200
TGAAGGAACG	CCTAAGACGG	CCGCAGCAAG	TATGTTGCCT	TGTTCTTGGC	TCTATCAAGA	4260
AATTGGTGCA	CAGTTAGTCA	AACAACACTC	ACCAGAGCCA	CTGTACCAAC	GTTGGATTGA	4320
AACCTATGCT	GGGGAGGAAG	CGTACCAACA	CGTCCAAGAG	GAACGCCAAC	TGTTAGACCA	4380
ATTATATGAA	GAAAGTTCAC	CACAAGAACA	AGCAGCTATG	ATCaCCGCTT	TTGTTATCAG	4440
CaGTGAGATG	GAATACGCTT	TTTGGGAAAT	GGCCTATACA	CACGAAACCT	GGATTGGTTA	4500
ATACAAAAGC	AAGTCGTTTA	GTGCTCGTAA	CTAACGACTT	GTTTTTGTAT	TAGAAAGCAG	4560
TAATTGTCAG	TTGCTTTTTT	GACAGGGCGT	GATAAGATAT	TCAGTGACGC	AAAATCTGTT	4620
GTAACGCT	CAGTAATAAC	AGAAACGCCT	TGCTAAGAAA	GGAAGTAGA	AGCCTCATGA	4680
ATTATCCAGA	ACCGATTGCA	AAATTGATTG	AAAGTTATAT	GAAATTACCA	GGAATTGGTC	4740
AAAAACAGC	GACACGGCTT	GCTTTTTATA	CAATTGATAT	GAAAGAAGAA	GATGCTAATG	4800
CTTTTGCAAA	AGCGCTAATT	AGTgTCAAGA	GAGaTCTCCA	TTTCTGTAGT	aTTTGCGGAA	4860
ATATTACTGA	AGAAGATCCT	TGTGAAATTT	GCCAAGATAA	AAATCGTGAC	CGTAGTATTA	4920
TTTGTAGTAGT	GGAAGAACCA	AAAGACGTTA	TGTCAATGGA	AAAAATGCGT	GAATATCAAG	4980
GCTTGTATCA	CGTACTCCAC	GGGGTGCTTT	CACCGATGGA	AGGAACAGGA	CCAGAAGATA	5040
TTAATATTGC	CTCTTTGATC	AAACGATTGC	ACGATGATGA	AGTGAAAGAG	GTCATTATAG	5100
CAACCAACGC	TACCACTGAA	GGAGAAGCGA	CGGCGATGTA	TTTATCACGG	CTCATTAAAC	5160

CAGCTGGAAT CACCGTTACT CGTTTGGCAC -ACGGCTTGTC TGTCCGCAGC GATATTGAAT	5220
ATGCaGACgA AATAACCTTG CTA AAAAgCAG TCGAAGGACG TCGAGAAATT TAAAAGAAAC	5280
ATCGTTGGAA AAAGAGGTTG AAATCGAAGT AGGTCGTTTC TGTGAGAAAA TGTCGCAGTT	5340
CAAGCTTTwC CGTACTTCTT TTTCAGTTTC TTTTTTTCAG CGGATGCATC ACTTTAAATA	5400
GGACATAAGG ATTTCAATTA GACTGATAGA GCAAGATGAA GTATAATAGA TACAAGTAAA	5460
AATTGATTGG AGTTTTCGCA CGTGCAGGgT ATTTTATCA CATTTGAAGG TCCAGATGGT	5520
GCTGGtAAAA CCAGTGTTTT GAAGGAAGTT TCTGAAnATT AGCTAAAGAG TCAAAACGAA	5580
AAATTGTCAC TACAAGAGAA CCAGGCGGCA TACCGATTGC CGAAAAGATT CGAACAGTTA	5640
TTT TAGATCC AAGAnATGAC AGAATGGATG AGCGAACAGA AGCTTTACTT TACGCAGCTG	5700
cTCGTCGTCA ACATTTAGTG GAAAAAATAT TGCCAGcTTT GGAAGCGGGT CATTTAGTCC	5760
TATGTGATCG TTTGTTGAC AGCTCATTAG CCTACCAAGG CGCTGGACGT nGAATTGGGG	5820
ATGGCCCCCA TGCCTCAATT AATGCGGTTG CGGCA	5855

## (2) INFORMATION FOR SEQ ID NO: 106:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8519 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

GTTTCATAAC TGTACCAATT GGAAAACTG ACATGGAACG GCGTAACTCT TTTTCTCTCA	60
TCTCCATCCT CCTCTCTTAG TTATTTCTTA GAATAACAAT AGAATGACAA ATTGTCAAGA	120
ACTTATGTTA GATTTTCTAA CATAAAACAA AAAACCTGAT AAATATGGAC TTTTCCATCT	180
TTTCACGGCT GATTTTCTCC TACTAATCAA AAAAATAGAA ATGATTTGTT TTTACAAATC	240
ATTTCTAATC AACTTACCAT TTTTATTCT TTTTCGGATT CAGCCAACCA TTTAGACTTT	300
TTTTTCGCAA AACCTTTCAC GATATACCGA TTTTCATTTT CAGCATATTC TTCTGAAATC	360
ATTAGTGTC TCCGGCGAAG TTCACTTAAT AGTTGTCCGT CAGTTGGGCT AATTTCTAAT	420
TGATAAGGCT CGAGTAATTC CATCATTTGG GCACGGATAG CTTGAACCAG TCGTTCTTTA	480
TCTTCAGAAA TTTTAGTTGA GATCAATACA TTGGGAAATA AGGTTGGGAC AAATTCTTTA	540
CTGTCCACTT GATCTCTTTT ATTATAAACC GTCAAGCACG GAATATTTTC CAACGCTAAT	600
TCTTTCATCA ATGTTTGAAC TGTGCGTTCA TGTTGTAAAC GATCTGGCGC ACTTGCATCC	660
ACCACATGGA GCAGTAAATC CATCGTCCGA CTTTCTTCCA ATGTCGACTG AAAGGCTTCA	720
ATCAGCTGTG TTGGTAAATC TTGAATGAAG CCAACCGTAT CGGTCAGCGT AACTTCCATG	780
CCTTGAGGCA ATTGCCATTT TTTGGTTAAA GGATCTAACG TTGCAAACAA TTGATCCTCT	840
GAATAAGTTC CTGCTGTTGT CAACATGTTG AGAATAGTTG ATTTGCCAGC ATTGGTATAA	900

CCAATCAAGC CAATTTGAAA TAGTTCTGAA TTTTGTCTTT TCTGGCGGCT TCGTTCACGA	960
TGAGCAGTGA CTTCTTTTAA CTCACGACGG ATGCCTAAGA TTTTATTCCG AATGTGGCGG	1020
CGGTCCGACT CTAATTTTCGT TTCACCAGGC CCGCGTGTCC CAATACCTCC TCCAAGTCGG	1080
GACAACTGTT TTCCTTGGCC AGCCAAGCGA GGCAATAAAT AATCCAATTG TGCTAATTCA	1140
ACTTGTA ACT TTCCTTCTTT GGAGCGCGCA CGCAAAGCAA AGATATCCAA AATTAATTGT	1200
ACACGATCAA TTA CTCTGAC AGCTAACGCT TCAGATAACA GTTGAtTTG ACGCGGTGTG	1260
AGCTCATGAT TAAAAATCAC AAGATCCGCT TCATAGGCAT CTGTAAAtTG AATTAATTCT	1320
TCCAATTTCC CCTTACCGAT TACGGTTTGT CGATCGACTT GTGGTCGTTT TTGCGTTAAA	1380
ACAAACACCA CTTCTCCATT AGCGGTTTCC GTTAAATTCG CCAGTTCTTC CATTGATGCA	1440
GAAAATGTCT GATAATTTTn TTCCGTTTCA ACTCCTACTA GAATCACTTT TTCATGAGTG	1500
GTCATCGCCT ACTCACTCTC CTTCTGTTGT AGCCATTCCG CTACTTCTTT TTCCAAC TTG	1560
GGcAAATCTG TTGGcTGCTG AACCAAATTC CACCAATGCG CAGCCATTCTG ATTTCTAAAC	1620
CAAGTCAATT GACGTTTGGC ATATCGCCGC GATTGCTGTT TGACTGTTTC AACCGCCATC	1680
TCTAATGATT GTTCACCAGA AAAATACGGG AAAAATTCTT TATAGCCAAT ACCTTGAGCT	1740
GCTTGTGCAT GTGGTTGCTG AAACATCTGT TTTGCTTCTT CCAACAGCCC TTCTGTCATC	1800
ATTTGATCCA CCCGCTGATT AATACGCTCA TACAATAAAG CGCGATCTGT TTCCAAACCT	1860
AACAAATAAT AATCATAGAG ACGAGCTGGT TTTTCCTTCG GGGTCAAAAT ACTATAACCG	1920
GTTTTATCAA AGACCTCTAA CGCACGAATC ACTTTTTTGC GATTATTAAA ATGGATGCTG	1980
TTAGCCGCTA GAGGATCCTT TTGTTGTAAT AACAGCCATA AGGCTTGATT CCCTTTTCT	2040
TCCGCAAATA AATTGTACGT TTCTCTAATT TCAGGCGAGT CATCTATCTC ACGACTTCCT	2100
AGTTGAAAT CATAAAGGAG TGATTGAATA TACAAACCTG TTCCTCCAAC GATAATTGGT	2160
AGTTTCCCTT TTTCAGTAAT TTCTTTGATT TTTTGACGAC CCTCTTCTG AAAATCAGCG	2220
GCAGAATATG TTTCGCTGAC TTCACGGCAA TCAATTAAAT AGTGAGGAAT TCCTTCTTTT	2280
TCTGTTTCTG TGACTTTAGC GGTGCCAATG TCTAACGAGC GATAAACTTG CATGGAGTCA	2340
CCGCTAATAA TTTCACCATT AAATTTTTTT GCTAAAGCAA TACTTAATGC TGTTTTCCCC	2400
ACCGCAGTGG GTCCTACAAT AACTAATACT TTCTCCAATT ATGTTTCTCC TTCTTCTGTC	2460
CAATTTTGAC GAGCAAATTT TGCTTCCTCA GGAAATCAG TATGAACACC CGCCAATTTT	2520
AGCTGAAAC AATTTTCCAT TTCTTCCATC CGATTAACAG TCCACGGCCG CAGCGGTTTG	2580
CCAATTGTGA TAAGTTCTTT TTCATGAGCC AACACCCAAG ACATTTTGG ATGTAAACTA	2640
TCGACAAGCG CGAATGCCGG TCCCGCTTGA CTAAATCCT CTGCAGATTC ATAAATAAAG	2700
GCAATTTCTG TCTTAGGATC AGCTTTTTTT AATTTAACTA AACTTGGAAA ATAAAACTG	2760
GAATATAAAT AACGAAACGG CCAGTTTTTT TGCTTCATTT GTTGGACGAT TTTTTTCTCA	2820
ATTCCTTCAT AATGGATAAT ATCGGTTTTA AGTTCAATAT TCAAAAAACC ATTGAATTGT	2880



TCTTCTACAA GACAATTTAA AACATCTTCC AATGTTGGAA CACATTGAAC AGAAGGATTT	2940
TTCTGAAACC AGCTTCCAGC ATCTAGTTGT TTTAATTCGT TTAATGTTAA CTGTTGGATT	3000
TCACCATGAC CATCTGTCGT ACGATCCACT GTTTCATCAT GCATAACGAT TAAATAACCA	3060
TCTTTAGATA GGTGAACGTC TAATTCAATT CCTTCAGCAC CAACACGAAC AGCCTCTCTG	3120
AAAGCAATAC ACGTATTTTC AGGATGGGTG CCTTTACTTC CTCGATGTGC AATAATATCT	3180
GTCATGCTCT CACTCCAATC TTGTTTATTG TACCATTGGA GTGAGGTGGG TTCAAATAAT	3240
GTAGAAAAAA GAAGCCTCTC ACAAATGAGG GGCTTCTTTT TAAGATATAA CAACTAACGC	3300
ATTCCTGGAG CATTTTTTTC TGTAATATCA AAATTTACCG AATCAATAAC TTTGTTAGTA	3360
TTACTATCAA TAATCACAAA GTTCAAACCT TTTTATTTT TAGACATTTT TGTATCATTT	3420
ATTTTACAG AAGCAAATGT TTGATTTTCA TAATTAGTAT AACTTGAGAA TACATCATAA	3480
TTTAAAGCTT CAATTTTACC AATTAGTCTC GATAATTCTG CATTTGTTTC AAATGTTTTA	3540
TTATCTTTTG AAATCACTGC ACTTAAAGTC CACGCATTGT TACCTTCGAA TTTTACGTCT	3600
ACCCCTAAGT CTTTTAAAAG CGACTCTTCC GCTTGTGAAA GATAGACTTG ATTACCATAA	3660
CCACTTAATA AAATAGTATA TTGCGTATCA ATTAATTTTT GTAAATAATC ATTAAGATTA	3720
AAAATATTCA TCAACTCTGG GTTACCATT CTATATATAT TGTCCCAATC TTTTCTGAA	3780
ACACTTCGCA CAGACTCCAA GCCGTAATAT TTTGCAACAG AACGATTAAC TTGTGCATCT	3840
GAATTAGTTT TTACATGCGA TAAATGATTA GAATAAGCTG AATATTTTGT TGTATTACTA	3900
TATGATATAT CTGGAAACAC AGGATTGATA TTTCTTTTT TCTCTTGCTC AACTAAATAA	3960
CTATATTGTT TTGTCAAATC TAATTTTGAC AAGAATACAT CACTTACACC AACCATTACG	4020
TTCATGAAGA ATGTACATAC CAAAATCACA TACAAAACAC TGTAATTGG AGAAAACCTG	4080
ATATTTACTT TATAATTTGG CAGTGACATA GAAAAGCGA TAATCAAGAA TAATACGGCT	4140
CCATAATATG AACGTCCGTA ATCTAATGCT GCTGGAGAGA TGGCTAAAAC TAGCATAGTA	4200
GCCATTCCAG AAATAAAATA TACAACACTT AGGTATACTC TTTTTTTATC TTGACTAATA	4260
TAAGCATATA AAACAAGCAA CAAGATCAGT ATCAATAAAA ATAATGATAA GTGTTCTTTC	4320
ATTTGCTCAA ATATTGTAAT AATTCCTGTC ATTAATTTTC TTGGTAAAGA CCAAGTACTC	4380
CTAGCAAAAT AAGTAGCGCG AATCTTATTT CCTGGTGCTA AAGCCATTAA CGCTAAACCT	4440
ATAGTATTAC CAGCAACACC CGTAAACATC CATAAAGAAA GTTTTCGTTT TGTATAAAAC	4500
TGCCAGCCAA GATACCTTAA AACGATAAGC AAAGCTCCTC CAGAGGTATT TTCATTACAC	4560
CAACCAGCCA AAATTCCTAA AGCAAACATC AAAAAATCA TTAATTTTGT TTTCGTAAAG	4620
ACGTGTTCTT CAATGATTTT TCTATGATAC ACATAAAGGA ACGAAAGAAT AATTAACTT	4680
CCCCAAAGAT AATTAGCTGA GCCTGTCTCC CACAAAACA CTTCTCCAAA GGCTGGGGTA	4740
AACAACCAAA ATAGCACTTG AATAATAATA AATTTAAATG AATTATACTT TTTATCTTGC	4800
TTTTGTAAAG TAAGTCTATA AACTAAATAT GTTACTATTG TGTATGCCAA AGCATTAATA	4860

ACATTAAAAA	TCCCTTTTGG	TAACAATAAA	AATATTCTTG	QAATTACATG	GACCACTGAA	4920
CGTCCTGTCC	ATGTCATGTA	TTGATTGTAT	TCATCTATAA	GTATCGTCAT	CCAACCTTTTC	4980
GTTTTATAAA	GGTATTCATA	GTCATCTGCA	AGTAATGGTG	TTAAATTATT	TAAAGTAAGC	5040
ATCACCAAAA	ATGTAATAAT	AATTGAGCAT	ATAAATATCT	TCTTTTTTATT	CAATTTAACT	5100
TtCATCTCTG	CTTACCTTCC	GCTTTTAGTT	CATACTTAAC	ATATTTATCC	AAACCAGATT	5160
GAATACTATA	TTTCGGTTCA	TAACCAATGG	CACGTAATTT	rGATATATCC	GCTAAGGAAT	5220
CTTkGATGTC	CCCAGCTCGC	GCTTCTTTGT	ATTCAACTGG	CAAAGTGACT	TTCATAATAT	5280
CATTTAAAGA	AGAAATTAAC	TCATTTAAAT	CGGTTGCTTT	CCCAGTCCCG	ACATTATACA	5340
CTTCCCAAG	TGATTGGTCA	GAATGAGCGA	CAAGATTCAA	TGCTTGAACC	ACGTCCTCAA	5400
CAAAGACAAA	ATCACGAGAT	TGTTTTCCAT	CGCCAAAGAT	ATTGAATGTC	ACTTCTTGAT	5460
TGGCTAATAA	GCGTTTGTA	GAGTCCATCA	CAATCGAAAT	AACGCCAGAA	TAAGGCGAAT	5520
TAGGGTTTTG	GTTTGGCCCA	TAAACGTTAA	AGAAGCGAAC	AGCGCTTGTT	GGAACATCAT	5580
ATAAATGACA	ATAGTTTAAC	ACATATTTTT	CAGAAGCAAA	TTTATCCACA	GCATAAGGCG	5640
TTAACGGACG	AATTACTGAT	TCTTCCTGTT	TTGGTAAAGT	TGGCTCATCG	CCGTAAACCG	5700
CTGCAGAAGA	AGCAAACACC	AAACGTTTTA	AATCTTTCTG	GTATTTCCGA	ATTAATTCCA	5760
GTAATTGTAA	CACACTCTCA	AAATTCACTT	GGTGTGTTTC	CACTGGACGT	GCTACTGAGT	5820
CTGCTACACT	AGCAATCGCG	GCTAAGTGAA	AAATGTAATC	GAACTGATAC	TCTTGTAATA	5880
CTTTTTCCAT	TAATTGTTGA	TCAGTCACGC	TACCTTCAAT	GAACGTGATG	TTTTCAGATG	5940
CGTTTAAATT	TTCAAGTTGG	CCCATTGATA	AATCATCAAT	GACCACAACT	TGATTGTCTT	6000
TACTATaAta	ATTTGCTAAG	GTTGAGCCGA	TAAAGCCGGC	GCCGCCTGTG	ATTAAAAAGT	6060
TCCTCAAGTT	TCCCACTCCA	ATTAAAAATA	TACATTAATT	TTACCATTTT	TAGGTGTCAA	6120
AGACTACAGT	TTGTTTTTTC	AATAAAAAAG	GAGAGAAAAT	AACTTTCTCA	ACCTTTTACT	6180
AAATATTTTA	AATCTTTTGC	GTTATTTACA	AAGCTCTTAT	GAATGAATAA	GAAATAGATA	6240
ATATACACAA	TTGCATAAGA	AATGAAACTA	TAAATACTAC	CGCCAAACCA	AAGATTACTA	6300
AAGATAAACA	ATAATGTAAC	CAACGTTTCC	AATGCAGTGC	TATTTCCCTAT	TTAATATTC	6360
ATATCTTTGC	ATAGAAAAAT	TTCCGCTAAA	TTACAACGAA	ATGCTAAACT	TACTAATATC	6420
AATCCAACAG	TTAAATTCAA	GTTCCCAACA	ACGAAAATCA	CAAACAACGA	TAAGATTAAC	6480
GACAAGGCAA	GTGTmAtaC	GtTAACAAAT	AAAATTGTTT	TTTCTTTTCT	CAACGTTTTT	6540
AAATACGTGT	TAATTAGTAA	AGACATCCGG	CCTTCATAAA	TAATATAGG	AAATAGAATT	6600
CCCATAAATT	TCAAACCTC	GCTATATTCG	GGTAACCACA	TACCTAAAAC	AAATTTGACG	6660
GGTACATATA	ACAGCAGAAT	TGCATACGTC	AACGGAACAA	AAACACCTCT	TAATGTGACA	6720
AATAAGACG	GcAATCGCTC	TTTATTTGTC	CGTCTTAACA	ACGGAACAT	TACAATTCCA	6780
ACAGCATTAA	TAAAAGTTAA	AAACATATTT	GATAAaCTTA	AAGTAAAgGm	CAaTTTACCA	6840

683

AAGGTTTCGA TTGTCCAACG CTGTGAACA AAAAAGCGAA TGGTTCCCAT TATTnGCnTA	6900
CTTGCTATAC TACTGAGCAT TAnTnACTA CCTATGTTAA TATTATCAaC TATTTCAGGA	6960
GTAAC TTCCT TCAAnCTCAT CATGTTTACA CGCaACATGT CTTTAACTCT AACATTCCC	7020
CAAACCGTAA TAATCAGTTT AGAAATATA TCCATATGA TTAGCCAATA AAAATCTCGT	7080
CCTCCTAATG CAAAGTAAAT TSCATGAAC AGGACATATA AATAGCGATC ATTTCTGGAT	7140
AATTGAGCAT ATTCCTTAAT ACGATTGTG TACTGAAAA CATACAGAAT AACGTTTTT	7200
GCAATTGTAA TTAGTAAAC AATGGCCGTC AATATCAAAA TAATCCCTTG ATATTCTTTC	7260
GGCATGAAAA AGTAAGCCCA AGAATGACA AAGCAGAAA TACAACATTC AAAATCGCT	7320
AAATACCAGA ATTGAGAACC GAGATTCTT TTATCTAAAT TTTCGTATTC TTCTCCCCCA	7380
ATTTTTAAAT AGATACCATC AATCCAGCCT AATGGAGAA ATCCAACATA TGAAGAATAA	7440
AAAACATAAA GTTGCCAATA ACTGTATTCT GTAACGCTA ACAACTTAGG GACAAATAAA	7500
TTTAAGAGAA CCGAATACC CAACGTTGCA AAATTEGCAG CCACAGTATA ATAAAGGTTT	7560
TTAGCCACCA TTTTGCTTT TGAATCATA GCATTATGCT CACTTTCTAA TTTTATATT	7620
CATGATATGG TAGATATCTG TGCTAAATA AACTTGTA TGAATCTAA AAAAGGCACA	7680
TAGAATCGGG ATTATTTTTT TTCTTTCTAA TAAAGAACT CGCATATTCA ACTCTTTAAT	7740
CAAATTTGAA AGAATTTGT CTTTATTTTCA AATTTCTTCA ACATGTGGTT GATTATTTAA	7800
ATAATTTCTT GCAGCCTCTG TACGCTGTAA TCTCCGGCGT AAATATTCA ATCTTCTCTC	7860
AAAGTGATT TGTTCCTTAG CTAATCAAT TTCTTTTTTA AAGAACTTT GTCCATGTTT	7920
ACGCCATCTA ATTAATTTCT CCTATCAAT ATAGAGCCCA TCACTTAATA AAGAACTTCC	7980
CCACATCGCC ATATCATGTA CAGGAATTTT CATTTGAGAT GCGTATAGAT TAACATTGGC	8040
AATGAAGCTC TTTCTAATG CATATACACA TCCCGGCGA CGTAAAAATA AATTGTTCTC	8100
ATTAAATATA ATCTGTTCTG CTTTTGATT AGGCAATTTA ACTGAATCAA TTTTTTTAA	8160
TTCCTCTGCT AATCCCCCAG GTTCTACTAA TCCAGTATAA TCACTTACTA AACTTTAAC	8220
ATCAGTGTG GCAAAGCTT CACACATCAC TTCAATTTA GATGCATCCC AATATCGTC	8280
TTGATCACAC GTAAAAACAA TTTCTTTAGT AGCAGCATTC AATAGATTAA AGAAATTTT	8340
GCGCCAACCT AAATTTTCCG KGTTCATTtG CAAATGCCAC GTTtCACCTA AwTTAwTTTc	8400
AGAAaTAAAG TtCTCAATtA TTTtAACTGT rGAATCyGTA GrGCAGTCAT CCCAAATTAA	8460
TACTTCATCA ATTTTTTTTCG ACTGAATTTT TTAGGATCC AACTGTTCTA AGATATTTT	8519

## (2) INFORMATION FOR SEQ ID NO: 107:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1508 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

TTTTGGGGGG ATAATCAAAT TCCCCTGnGT TCCGCCACAA GCCACACAAT GATGCCcATT	60
CTGGTCGTAC ACCAAACCGT TCTAACAATT GAATTTCAAA AATATTTGTA ATGATTTCCg	120
GATCTTTTTG CTCaTTTAAA GCGGATAATG CCATGTGGGT AAATTGAAAT AAGTTCGGAT	180
CATAGTGCTG ATCTTCAATT GCTGCATCGA CTAAGTTTAA AATATATGTT CCATATGCGT	240
TAATAAATAT GTCTTGTTGA ATGTTTCGAA AAGGTTGAAC TTCTTTGCTG CTATTGAGAA	300
AAGATAAGCC TTCTTCTCTA AAGTTTCCGA TGTAGACGGC TTCTGTAAAT GGCAGAATCG	360
CTGGCAAAAG AGGATTATTT TTTGATGCG CACCTTTCAC AAAAAACATC AATTTTCCAT	420
AGGATTCTGT AAAAATTTTG ACGAGTTTAT CCTTCTCTTT AAAATCCTTC GTAAATAAAA	480
TAATGCCTTT GGTCTCCCTT AATGCCATTC AATTCAC TTCATTCTTAT GTAAATTAGT	540
GTAGTGTTTG -GAAAGTTCGT TTATACACTC GCTTCCCAAA CACTACTTTC TTAATATTCT	600
TCTTTGCGAT AACCAAAATC TTGTAAATAG ATTTTTTTGT CACGCCAGTC TTTTGGACT	660
TTTACCACA GCTCCAAATA CACTTTGTCA TCTAATAAAT ATTCAATATC TTGACGCGCC	720
TTGGTACCAA TTGTTTCAG CATTTTTCCG CCTTTGCCAA TAATGATACC TTTTGGACTG	780
TCCCGTTCCA CAATAATGGT TGCTTGACA TGGACTTTAT CGTTTTCATC CCGTTTCATT	840
GAGTCCACAA CAACAGCGAC AGAATGAGGa ATCTCATCTC TTGTAAAAG TAACACCTTT	900
TCACGAATTA GTTCAGAAAC GATAAAATAT tCTGGATGAT CCGTAATTTG GTCaTCaGGG	960
AAATATTGTG GTCCTTCTGG CATTGCTcCA ACGAGAACAT CCATCAAATG CTsCAAAATT	1020
ATTCCTTCA GTGGCAGAAA TGGGAATAAT TTCTTTAAAG TCCATTGTGTT TCGAATAGTC	1080
TTCAATAATA CTCAATAAAT CATCTGGATG GACTTTGTCA ATTTTATTAA TAATCAAGTA	1140
GACTGGACTA TTATTATTTT TCAATCGTTC AATAATAAAG TCGTCACCTT TCCCGCGTTT	1200
TTGATCAGCA CTTACCATAA ATAAAGTAGC ATCTACTTCA CGCATCGCAT TGTAGGCACT	1260
TTCCACCATG AAATCGCCTA AGCGGTGTTT AGGTTTGTGA ATCCCTGGAG TATCGATAAA	1320
AACAATTTGT GCTTCAGGAA TTGTATAAAT TCCTTGAATC TTATTACGAG TTGTTTGTGC	1380
TTTGCTACTC AtAATTGCAA TTTTGTGGCC GACAATTCTGA TTTAACAATG TTGATTTnCC	1440
AACAnTTGGT CTGCCGACAA TGGGCTACGA AACCAGATTT AAGGTTcAGT TGTCATGGAA	1500
TnCGCCGC	1508

## (2) INFORMATION FOR SEQ ID NO: 108:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5918 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

ATGAAGTATT CAAAAGGACG GTACGTCCTT GTAGTCTTAG TAATGGTCTT GATTGCTTGG 60

TTAATTTT	TTTTGTCAGG	TTTAGCGAAC	GGTTTAGCTC	AAGGAAATCG	TTTAGCCGTA	120
GATCAATGGC	AAGCGAATCA	AGTGGTTTTA	TCAAAAGAGG	CCAATAGTAA	TTTAAATGTA	180
TCAGTGTTAG	ATGAAAACGT	GAAAGAAACG	ATTTTCAGGAG	GCAAAATTGC	ACCGATTGGT	240
CAACAATCTT	TAGCCATCCG	CCCAGCAGAT	GATAAAAAGG	CTGAATTAAC	GAATGTTAGC	300
TTATTTGGGA	TTGAAAAAGA	AAGTTTTTTA	ATGCCAAAAG	TGATTGAAGG	AAACGCATTT	360
ACTGATAAAA	ATCAAGTGAT	TGCTTCaGAA	ACATTGAAAA	ATCAAGGATT	TAAAATTGGC	420
GACAAATTAA	CTGCAGGGAA	ATATGATGAG	CAATTAGAAA	TTGTCGGCTT	TATTTCTAAA	480
AGTAGCTATA	ACATTGTGCC	AGTTATTTAC	ACTTCTCTGG	ATACTTGGCG	GTCAATTAAA	540
TATGGTGACA	ATCCAGCAAT	GGCCAAAATG	GTTAATGGTT	TCATTGTTTCG	CAGCAAGGAC	600
AACACGGAAG	TTAAAACGAC	TAATAAAGAC	AGCCAAGTTC	TTCAATTTTC	AGATTTTATT	660
GAAAACTGC	CAGGATACAG	CGCTCAAAAC	TTGACCTTGG	ATGGCATGAT	TTATTTCTTG	720
ATTGTGATTG	CAGCGTTCAT	TATCGGTATT	TTTATCTTTG	TGATGACTTT	ACAGAAAACA	780
GCAATGTTTCG	GCGTCTTAA	AGTTCAAGGG	GTGCCAACTA	GTTTCTTAGC	AAAGGCAGTC	840
ATGCTACAAA	CCGCTTTTATT	GGCTGTTCTA	GGAGTGGCGA	TTGGTCTTGC	TTTAACGGGA	900
ATTACCGTGC	TCTTTTTTACC	AGAAGCGATG	CCTTATGCAA	CTAATGGACC	ACGCATGATT	960
TTGTTTAGTG	TACTATTGAT	TTTATCTGCA	TTAATTGGCG	GGGCATTTTC	AATTCGAACG	1020
ATTGCTAAAA	TTGATCCGTT	AATCGCGATT	GGAGGTTAGA	GACATGGCAA	ATGTTTTAGA	1080
AATGAAAAAT	ATTTATAAAA	AATATGGTGA	AAAACATACA	GAAGTGATCG	CGTTAAAAGA	1140
ATTATCTTTT	GCGGTTTCAGC	CAGGTGAATT	TGTCGCAGTA	ATTGGTCCTT	CTGGTCTCGG	1200
TAAAAGTACC	TTTTTAACGA	TTGCAGCTGG	TTTACAGGCA	CCGACAAGTG	GCGAAGTCAT	1260
TGTTGGCGGG	CAATCACTAA	ATAAGTTAAC	GAAGAAGCAA	CGTTTGGCGC	AACGTTTTCA	1320
AAAAATCGGC	TTTATTTTAC	AAAGCTCTAA	TTTGGTGCCA	TTTTTAACAG	TGGAAGATCA	1380
ATTTCACTTA	ATTGAAAAAG	TTGATAAGTC	ACGTAAAAAT	AGTGAATTAA	AAGAGCAATT	1440
GTTAGAGACG	TTGGGTTTAA	AAGAATTACG	AAATAGTTAT	CCTCGTGATC	TCTCTGGTGG	1500
GGaAAGACAA	CGGGTAGCCA	TTGCGTGTGC	ATTGTATCaT	GAGCCAGACG	TAATTTTGGC	1560
AGATGAACCA	ACGGCTAGTT	TAGATACAGA	AAAGGCGTTT	GATGTTGTCA	AATTACTAGC	1620
GAAAGAGGCC	AAAGAGAAAG	ATAAAGGGAT	TATCATGGTG	ACACATGATG	AGCGTTTGTT	1680
GAAATACTGT	GATCGAGTGG	TTCGTATTTCG	TGATGGTGAA	TTGACAGAGT	AACAAAACAA	1740
ATAAAAGaGA	ACATGATCGG	ATTGCTGGAG	CAATCCGATC	ATGTTCTCTT	ATTTCTTTTC	1800
AATACAGAAA	AGAACAGGTG	GCTGATTTTT	TTGATTGATA	AACTCATAAC	GTAAGACACT	1860
GTAAACATCT	TGTGGTAAGG	CTTGCCTGTA	ATTCGTGACC	AAATCAAGTT	CCTCTGAACC	1920
ACCTTCGTGG	CCATAATAAA	CGACTAAAAT	AATGCGACTT	CCTTTTACCA	AATGAGGCAG	1980
TAACCCATCT	AACGCCTGTT	TAGTCGTAGT	GGGTTTTGTA	ATAATTTCCCT	TATTACTTTT	2040

AGGTAAATAG CCTAAGTTAA AAATAGCGGC AGTGATTTCT GTTCTTCAT CTAGGACAGC	2100
ATGAAGATGT TCATGTCCTA ATGGAAATAA CGTAGTTTGT GGGAGTAAAT TAAGTTCTGT	2160
TAACTTTTGT TCTGTATTGA TCAACGCTTC TTTTGGATA TCAAAGGCGA AAACCTCACC	2220
AGAAGGGCCA ACCAGCTCTG CTAAGAATGC TGTATCATGA CCATTTCCTA TTGTTGCATC	2280
AATGACGAAG TCGCCAGGTT CCACCACTTC TTTTAAAAGT TGGTGACTAA AATGTAGGGC	2340
TGTTTGTAGC ATTAATTGAA AACTTCCTTT CCTGAAATGA CAGTATATTT TCCTTGATAA	2400
CTGTTGCGGC GTTTCATTTT CGCATCAATC GCATTGAGAA CTTCCCATTT TTTCAAGCTC	2460
CACATCGGTC CGATAATTGT TTCAAAGGGA GCATCGCCTG TCAAGCGATG AATAACAATT	2520
TCGGGTGGAA TCATTTCTAA CTGATCACAA ATCACTGAAA CATAGGCTTC TTTGCTCATC	2580
AATTGTAAGC GTCCTTCATT ATAATCCCGC ATCATTTTGG TATTGTTCAT TAAATGAAGT	2640
AAGTGCAGTT TAATCCCTTG GATATCTGAA TCCTGAATAG TTCGCCGCAC ATTTTCGCGC	2700
ATCATGGCAG GTGTTTCCCC AGGTAGCCCG TTAATTAAAT GTGTACAGAC CCGAATCCCG	2760
TGTTTCCGTA ATTTAGCCAC ACCATCTAGA TATGTTTGAT AATCATGAGC ACGATTAATG	2820
GCAGCGCTAG TTTCTCAAA CGTGGTTTGC AAACCCAACCT CTACCCATAA ATAAAAGCGT	2880
TCATTTAATT CGGCTAAATA GTTCACTACC TCATCAGGCA AACAATCAGG ACGGGTACCA	2940
ATGGATAAAC CAACGACACC TTTTTCATTG ACTACTTGTT CAAACCGATG ACGAATAACG	3000
TCAACGGGCG CATGTGTATT GGTGAAATTT TGGAAGTAAA CAATGTATTG ATCGACAGTT	3060
GGCCATTTTT GATGCATCAA TTGGATTTCT TTTTGAAATT GAAGCGGCAG GGGATCGCTA	3120
GGTGCAACAA TCATGTCGCC TGAGCCAGAA ACACTACAGA ACGTACAACC ACCTTTGGCA	3180
ACCGTCCCAT CACGATTGGG GCAATCAAAG CCACCATCAA TTGGCACTTT AAAAATCTTT	3240
TCACCAAACCT GTTGACGAAG TGCATAGTTC CAAGTGTGAT AGCGTTTATT TGGATCATCT	3300
GAATAGGGAA AAATTTGCAT AATCAAGACT TCTTTCTAGC GTTGTGCTT TGTAAATCGC	3360
CAAACAAAGC GTTTTTCTAA ATAAATAGGG TAACCTAATA ACAACCAAGC TAAACCAAGA	3420
CAAAAACCAC CTAAGATATC GCTAGGGAAA TGGACACCTA AATAGATTCTG ACTAACCCCA	3480
ATCATGAAAA TCCCAATACC TAGTAAATTT TGCACACAAA GGCGGACCGT TTTATTTTAA	3540
ATGAAAATTG GCAGAATCAA TAAATAGTG CCGTAGAACA AGGTACTTCC TACAGCATGT	3600
CCGCTAGGAA AACTATACGT ATGTTAGTT ACTAAATGCT CCAAAGTAGG ACGTTGGCGC	3660
ATAAAGAATA ATTTTAATAA AGAATTACCT ACGCCAGCAA TTAAGGCAGT ATTAATAAAG	3720
AGCCATAATG CTTGAGCATA ATATTTCCAA ACAATAAACA GAGCAATGAC AACGATTGCT	3780
AAGATGATAA TCGTCAATGA ATTGGCAAAT TTGGTGTACC AAATAAAAAA GTTATTTAGC	3840
TGTGGAAGAG GACGACGAAC AAAAGAAGTA ATCGTTTGAT CAAAGCCTTT GAGCCAGGTT	3900
GGATAAAAC GAACAACATA GCCTAAAAAC ATAAATACTA AAAGAAAACA ACTACCTGCA	3960
AATTGATAAT AAAGTTTGTT TTTTAAAAA AATCCTTTCA AAATTTAATC GCAACAATCC	4020

GTATTATACC ATCAAAAGGG TAAAAAATTA AAAGATAATT TCATTGCCTG ATTTGGAGGG	4080
GATACAGATA TTAAAAABAGT TCTGGCACCA CAATCTCTCA TTGTGGTGCC AGmACAGATG	4140
TTTTAACTGG CACATGTTTA TTGATTATCA TCATCAGGAG TATCAACGGT AATATCTCCG	4200
AATTCGCTAG AAAGGTTAAA CTCTGTATCC GCACGACCAT TCTTATAAGA TGTTGCTCT	4260
TTGTTTAAAA GTTGAACATC GCCTAATTTT GATTCCCCTT TGATTGAAAT GTTACGCAA	4320
ATTTCTCCAG AATTTACATA AATATCTCCA TGGACAAGAG CGACAGTTAA TTTTTTAGGC	4380
ACAGCAGGCA TGTCTAAAGA AAAATCACCA TTTTCACTTG AAAAGTTAGC TTCTCCTTTG	4440
AAATCATAGA CGCTAATATC GCCATCTGTC GTTGTGCTT GGATGTTTCC ACGAGCATTT	4500
TGCACTCGAA TATCACCAGA AGTAGTCGTA TTTTGTATT CATCGAAATA GGCTGAATCC	4560
AACGTTATAT CGCCGTCAGT GGTTCACCT GTTGCTGTT TAGCGGAAGT ATCCGCTGAA	4620
AGGTAGATAC CATCTGAAGA GTTAAGCGTC ATTTTTTTGG cTTTAACACC CGATAGACTA	4680
ACATAAGAAT TTGGtAGATT AGTTGTAAAT TCGTCTACGT TAAAAGCATT CAATGAACT	4740
GGTGAATGGG TTGAGCCATC AATCACTAAC TTTTTCACAT TATTAGGAAT CTGGACACTC	4800
ACGTTCCGGT TGAAGTCACT AAAAATACCA ATATTAAAGA ATGACAATTC GGATTGGTTT	4860
TGTGTACGAG TGGTATTGAT GGTGCTGTT ATTTTATTTT TTTCTTCTTT GACATCTAAG	4920
GTGCCTTCAG CTTTTGAATA AAAGGCATTA TTCGGTGATC GTAACCCCAT TTTGTTACTA	4980
TCAGTTTTAT AAATGGTATA AGGTCCTTTG CCTTTTAAGG ATAAGTGAAG TTCTTCGCCT	5040
GCTTGTATAT TTTTATAGT ATAACCTTCT TTAGCGGTG TGAAGTATT TTCCGCTTG	5100
TTGTAATAAT AGACGGGCC AGATCCACCA ATTAGCATCG CTAAGACAGC AATGGATAAG	5160
AGAATCCCTG TTGTTTTTTT CaTTAGATAT CACCTCGTAA AACAGCAATG TTCCATTGGA	5220
AATAACGCTT AAAGACATGT GCTGAAAATT TTAAAATTGG TAGAAACAGT AACCAACCAA	5280
TGATTGATGC ACCAAATAAA AAGAGACTCA TAGATAATTG GAAAAAGGAA GCGGCATTTG	5340
CTTGGGTAAT AACACTATAA ATaCCATAAA TTGGTGAAAA TAAAGTTGCA GCTCCGCAA	5400
TGmACAACC GACGTACATT AAGGCAATCC CTAAAATAAC CCAAACATG AAGrAAAAAT	5460
TTAAACAGAA AATACCAAGC ACTTGAAAA ATCGGGTTAG TGGCCGATGC TGTGGTCGTT	5520
GATAGTATGG AGAATCATT TCATAGTAA ATTCAGATTC TTCTTCATAG GGATGATCAT	5580
AAGCAGGTGT AGCCGCTGGA ATTTCTTGCC AGCCATTATT TTCAATTTTT CGTCCGTCA	5640
ATTCATGCC TTGTTCTTTC AAAATTTCTT CTGCGATAAC CCGTGGTTTC CCCAATGATT	5700
TTGCCACTTG TTCTTCTGTT TCGCCTTCCG CAACACGAGC ATCAAATAAC GCCTTGATT	5760
TCGCAAGAAT AATAGCTTGA TCTTTTGAT TTAATGCTTT TAAATAAATT TTCAGCTCAA	5820
TAATAAAATG TTCTTTATTC ATCCGCTCC CTCTTCTAC ACGATTCATT TCTCATCTTA	5880
AACTATAATT TCTTTTTGAG CGAATAACAA GAGGAATT	5918

(2) INFORMATION FOR SEQ ID NO: 109:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2350 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

CATACCCGCT CAATCATGAG GAGACTGGCT AAAAGTAATC CTAGCGTTTG CATCTTTTTA	60
TTTTTATTAA AGAAGTGCCC TGCTAACCAA CCTTGTCGCG CATGAATCAA TAACGAAAAA	120
ACGAACCATT GTGGATAGCC TGACAATAAA TCAATTAAAGC CACCGCTTAA AGCTCCGACT	180
GTCAAGCCAC CGATCGGACC AAAAAGTAAA CTAGCAATGT AAATCCCTGC ATCCGCCAAT	240
GTAACAAAGC CATTTGTTGC TGGGACGGGG ATAATAAAGA CCATTGACAC AGCCACTGTT	300
AAAGCGGTAA TTACTGCATA TAGACTAATT TTTCTTGTC A TTATCCGCCA CACCTCCTGC	360
TAGTAATCCT TCATGCCACA CACTTCCAAA AGCTGCATTT AGCTGGACCC CAGAAACAAT	420
GGCTTCATGA ACGTATGTTT TTGCTAAATC AACTGCTTCA ATTAACGGAT ATCCTTTAAC	480
TAAATTCGCA CCAATTGCAG AGGCAAAGGA ACAACCAGCG CCATCAATGG TGTGGGCCAT	540
TACTTTTTCT TTGACCAATT GAATATGTTG TTGTCCATCA TAAAAAAGT CACGTGCTTC	600
TTGACCAAGC CAACGATTAC CACCTTTAAT CACCACCGTT GCCGCTCCTA AAGAGTTTAA	660
CTCTTCCGCC GCTTGTGGA CTTCTTCGAC GGTGTTAAT TTTCTGCCAA TGAGTAATTC	720
TGCTTCTTTT AAATTAGGCG TTATAATCGT TGCGAATGGA AAAAGTTTTT CAATCAATAA	780
CTGCATGTAT CTTTTTTCAT AAAGTGAAGT CGTTTCTTTA AAAGCCAAGA CAGGATCCAA	840
CACCACCGGC AGACCTGCGT GCTTTTTCAA AAAGTTACA ATTTGTAACA TACTTTCTTC	900
ACTGTGAATC AAGCCAATCT TGATGCCGTC TAAACTCCCG GcAGGTATGG AAGCGAGTTG	960
CTCTTTccAC CAAAGAAgTC TCaAGGTCAT GAATCTGAAA ATGATTCTTG GGGGCTGTAG	1020
CAAGGCAAGT TAAAGTTGAT AAGCCAAAGA CTTGATGATT TTCAAACGTT TTTAAATCTG	1080
TTTGAATCCC GCCACCTGAA AAAGGATCGG AACCGCCAAT TGTTAAATTT ATTTTGTCA	1140
TTTGTTCGCG TCCTTTATAC ACTCAGTTAT CATACAAAGT TTTCTAAAAA GAAAAAACAG	1200
CCAATTGGCA GATTTTTAGC TCAGCCACTT TTTAAGTCTT TCGTTATCTT TTCCTTTCC	1260
AAAAAAGAG TGACCTCATC CGTCAATGAA GTCACCTTTT CTTTGGATTT AACACTTATT	1320
CGTCAATGGC TGTTACATCA ATGTTTAATT CAAACAATTG TAATGGAGAA ACTAAACTTG	1380
GTGCGTTGTT CATTGGGTCA ATTGCTTTGC CGTTTTTAGG GAAGGCAATC ACTTCACGAA	1440
TATTTTCTTC CCCAGCCAAT AGCATTGCTA AGCGGTCTAG CCCTAAAGCA ATCCCTCCAT	1500
GTGGCGGGAA GCCGTAATCT AATGCGTCTA GCAAGAAACC AAATTGGTCT TGTGCTTCTT	1560
CTTTTGTAAG ACCTAACGTT TCAAACATTT TTTCTTGTA TTTCTCGTGA TGGATCCGTA	1620
ATGAGCCGCC GCCTAATTCG TAGCCGTTTA AGACAACATC GTAAGCTTCA GCATACACAG	1680
AGGCTGGGTC TGTTGCTAAG CGGTCCACAT CCTCTGCTTT TGGTTGTGTA AATGGATGAT	1740



GAGCAGAAAC ATAGCGACCA GCTTCTTCAT CATATTCAAA CAACGGCCAA TCCACAATCC	1800
ACAAGAAATT GAATTTGAT TCATCAATTA AACCTAATTC TTTACCTAAA CGTGTACGTA	1860
CCGCACCTAA AGCAGCGGCT ACGATTTCAG GTTTATCGGC ACCAAACATT AAAATATCAC	1920
CGACTTCTGC ATTGGTAGCA GCAATTAATT CATCACTAAC TTCTGTTAAG AATTTAGCAA	1980
TCGGTCCTTT TAAGCCATCT TCTTCAACTT TTAACCAAGC TAAGCCTTTT GCGCCAAATT	2040
GACTCACATA CTTACCTAAG TTATCCATAT CTTTACGAGA ATACTTATCT GCGGCGCTT	2100
TTGCATTCAA AGCCTTCACA TGTCCACCAT TTTCCAACGC TGCTTGGAaA ACTTTGAAGT	2160
CTACGTCTtT CACTACCTCT GCAACATCAA TTAATTCCAT AGCAAAACGA GTGTCGGGTT	2220
TATCGyTACC ATAGCGTGCC ATTGCTTCGT CATAGCTCAT TCGTGGAAC GGAACAGArA	2280
TCTCAATTCC TTTTGTCTCT TTCATCACTT CAGCCAACAT ATTTTCTGTA TATGTTTGAA	2340
TTTCTTCTGG	2350

## (2) INFORMATION FOR SEQ ID NO: 110:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21591 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

TGAATTGGCC TCGATGACAG GTCTTCTACG AGGATACGTG CAACAATTAG ACGTCCCTGA	60
ACAGCGTGCA CTAACAGAAG AACTAGCCAA AATTTGCGAG TTAATCTACC ATGTCAATCC	120
AACGACCCGA ACAAATTA CAGTCACAGA GGACGAAATA GCCTGGCTTT TGGAACGTGT	180
CAATGCTATG AATGAACTGA CATAACGAAGA AAATCGTCCT TTTGTTTTAC CGATGGGCAC	240
AATCTGTTCC TCTTATGCAC ACATTTTAAG AGCCAAAGCC AAAGATATTG TTCGTTTACT	300
TTATCGCATG GACTATGGTG GTAAAAAAT TGATCCGCAG CTGTATGATG TAGTCAACTT	360
ACTCTCAGGT TACTTTTTCA TGTTAGCCTT GTACCTAAAT CAGCTCGAAA ATGGTGAAGA	420
AGTTCCCTTT GTTTCTCGCA ACTATTCTAT TTAAAAAGTA CACCGACCGC TAACGACGAA	480
TTGCTAGCGG TCGGTGTACT TTTATACAGG TAAGTACGCT TCTCCTTTTG CAAATAGTTT	540
AGCAGTACCA CCAATAATGA CTCGATTTTC TTTCACTTCA CAAGTTAAAA AGCCGCCTCT	600
TGGGGACACT TGATAGGCCG AAAGGGTCGT TTGATTGAGC CGTTTTCCCC AATAAGGAAT	660
TAAATTGGCG TGCGCAGAAC CACAACTGG GTCTTCGTTA ATTCTTAGTT TCGGGAAAAA	720
AGTTCGTGAA ACAAATCCA CCGAGTCTCC TGATGCCGTT ACAATGACTC CTACTCCTAA	780
ATCCAATGCT TTTAATGCAG AAAAATCTGG TGTGATTTTT GCCACTGTCT CTTCGTCTTT	840
TAAGACAAAA AATAAATCGC GACCCAAATA AGCTTCATAA ATTTTCGTGC CTAAAGCGGC	900
TTCATACTCT GGCAAAATTG GAATCCTTTC CGGTAGGATA TAAGGAAAAT CTAGATAATA	960

GTATTCCTCT	TTTTTGGTCA	CTGCGAGCGG	TCCACTTTGA	CTAGTAAAAT	GGAGTGTTTC	1020
TTCCGCAACG	GAATAATAGT	TAAACAACAC	AAAAGCCGTC	GCTAATGTAG	CATGTCCACA	1080
GAGATCGATT	TCTCGTTCCG	GCGTAAACCA	GCGCAGGGCG	TAACTTTGTC	CCTCCTTGAC	1140
CGTAAAGCA	GTCTCTGACA	AATTATTTTC	AATCGCAATG	TTTTGCATCA	CCGCTTCTGG	1200
CAACCATTTT	TCCAAAACAT	AAACAGCGGC	GGGATTGCCT	TTAAAAACTT	CCTCCGCGAA	1260
AGCATCGACA	ATATAATAGG	GATAACTCAT	TTTAATTCTC	TCCTTGATT	TTTTGATAAA	1320
TGTATTCCGC	GCAACACAAA	TCCTCAACTG	CTAAGCCAAC	CGCATCAAAA	ATCGTGCGGC	1380
CTTTCTGCTC	CGCCTTTAGT	TTAACTTTGC	CAGTGACTAA	TCGCCGAGA	GAACCTACAA	1440
TTTTTTCAGC	TGAAAAGGCA	CCTTCTGAAA	TCGGAATTAA	AAGATCCCCG	CTTTCTTTTA	1500
ATGCAGCCGC	ATACTCATCT	ACAAACACTT	GACTGTTTTT	AACAAGGTCG	CTTGTTAATT	1560
CTCTGGAAGT	CGGCGTGAAT	GTACCAATTG	CATTGATATG	GGCATTCCGC	AAAACGTCTT	1620
TTTCTTCTAA	AAATGCGTCC	TGACTGGAAG	TTAGTGTACA	GATAATTTCC	GCTTCCTGGA	1680
CGGCTTCTTG	GACAGAAGAA	CAACCAATAA	AGGTACATTC	GGGGTAATTG	TGCTGTTGCT	1740
TAGCGATTAA	TTTTCGTGCA	CGTTCTGCTG	TGCGGTCATA	TATAAAAACC	TTTCAATCG	1800
CCCGAACGCT	TTGTAGCGCT	GCGAGATGAG	ATGAAGCTTG	TTGACCTGCA	CCAATCAAAG	1860
CGATACTTTT	TGCCTTTGGA	GCTGCTAAAT	AATCGGTTGC	CAAAGCAGAA	ACAGCTGCCG	1920
TGCGAA\$CCA	CGTAATCGCA	TTGGCGTCCA	TCATCGCGAC	TGGTCTTCCA	GTAGAACTAT	1980
CAAACAATAA	AATTTGACCA	AGATGCGAGG	CCAAATTTTT	TTGTGGATTT	TCTGGA\$AAG	2040
CTGTAATGAT	TTTTGCCCCA	AAAAAGCGAT	TTTTTCCTAA	GTAAGCAGGC	ATTAATGCAA	2100
AAAGATTTTG	ATTTTTCCCA	TCTGGTAAAA	CCTCTACATG	GCGCTCACTT	TGCGAAATTT	2160
TTCTTTTTTG	CAACTCTGCA	AAGCAACGTT	TCATCACTTG	AATTGCTTCG	GAAAACGACA	2220
ATTGTTGAAT	AACCCCTCTCA	TAAGAAATAA	TGTCCATCTG	AATTCCTCCC	TTTGCGCTAT	2280
CATAAACCAG	ACAAACAATA	CACTCAATGC	TATAATTGTA	TGGAATACAA	TTTTATAGAA	2340
AGGAAGTTCC	CCTCATGCCG	ATTAATTCCT	ATGAGAACTA	TGTGTTGTCA	TGGCGCCCTG	2400
TTAAAGAACG	TTTAACAAGA	CCTTATTACC	AATCATTGGT	TCAACAATTA	GAAGCGGATA	2460
TCTTATCTGG	TGCGCTTCAA	AAAAATGTGA	AATTGCCTTC	GCAAAGAGAA	TTAGCTGATT	2520
ACTTGATCT	GAATTTACAG	ACGATTGGTC	AAGCCTATAA	ACACGCTATG	GAAAAGGCT	2580
TACTTTATAC	CAACATCGGT	TGCGGCACTT	TCGTTTCTCC	CAACGCGTTT	CACTCAATTA	2640
CAATCTCTAC	CAACCAAGTA	GCGGACCATC	TAATTGATTT	AGGGCTGGTT	AGTTCCTTTG	2700
ATATGTGTAA	TCAGCGGATT	TTACCTTTTA	TTGAATCCGT	TAGTAAAAAT	GCGGCCCTTA	2760
ACAGTTTGCT	AAATTATCGC	GATCCTTTAG	GTACGCACTT	TCAACGAGCA	ACCGCTGCCG	2820
AATGGCTTCA	GACACAAGGC	GTTCGGACCA	ATGCCGAAGA	AGTTGCCATT	GTATCTGGTG	2880
TCCAGAATGG	ACTGGCCGTG	ACGTTAGCCG	CCGCTTTTTT	TCCAGGTCAG	CGGATTGCCG	2940

TAGATCGATA CACGTATTCA AATTTTATTG AACTCGCCCA GCTTTATCAT TTAGAAATCG	3000
TCCCGATTGG CTATGATTCC GAGGGGATGG ACCCAGAGCA TTTACTGCAA GAATGCAAAA	3060
AGAAAAAAT ACATGGTATT TTCTTAATGC CCGCATGTAA TAACCCGATT GGTTTCCAAA	3120
TGTCTAGCGC CCGCCGAATG ACTTTAGCTG AAATCATTCA GCAAGAACAC TTATGGGTAA	3180
TTGAAGATGA CATTCATTCT TTTTAAACAA CTTATGCACA GCAAGCAGTT TTACCAACGT	3240
TTCAAGAACT CTTACCTCAA CAAACGATTT ACTTAGCGGG GATGACAAAA TTTGTCTGTA	3300
CAGGcTTACG CATTGCGTAC CTTGTTTTCC CTCCGCTTCT TCGCCAAGAA ATAGAGCGTG	3360
CGATTTTTTAA CATTAACTTT AAAACTTCTG GCTTTGATGC TGAAATTGTC ACCCAAGTCT	3420
TACGCTCACC AGTTGCAAAA GAATTAGTGA TAGAAAAATT GGCTTTAACA AAACAAGCCA	3480
ATGACTTATT TGATACCATT TTTGGGTTAG CTAGACCTAG CAATCCCTTG CCGTACTACC	3540
GAACATTCC CTTTCTACC GAAAAACGG CGCCGCAAAT TGAACAAGAA TTTCTCCAAA	3600
ATGGGGTACG TCTTTTTAC TCGAGCCGCT TTAGTGTCGA AAATCAGCCA GATGCTTTTT	3660
TACGAATTTT CTTGGCCTCC AATCAGCTTG AGGTATTGGC GAAAGGATTA ACCATTATCC	3720
AAGAACTCTT ACCGACATTA AACGAAAAGA AAGGACATTC CCTATGAATT CATTGCCAAT	3780
ATTTGAAAA GCAGTGGGCA CGATTACCCT GCTTTCCGAA GAAAAACAAA ATGCCGAATA	3840
TGAAGGCTTT TTATGTCAAG TTAATGAATC GCAACAACCT ATCCGAGTC GCTTAGCCAA	3900
AAAAACGCCG AAAAAAGAAG GCTATTTTGT AGCCTTCTGG GAAAAAATC AACAAATCA	3960
GAACGAAGCC TTTGATGCAA CGGAAGCCCC TGAAATGTGA GCAATTGTCA TTGCTGATCA	4020
AGAAAAGCAA GGTCTCTTTC TTCTGCCAAA AGAATGTTTA ATCCAACAAA AAATTCTCAA	4080
AACGCATCAA CAAAAAGGCA AAATGGCGGC CCGTTTTTAT CCTTCTTGGT GCCAAAACCT	4140
TAATCAAACC GCCAAAAAAA CACAAAAATG GCAACTAACT TACTTTACTG ATTTATCTAA	4200
ATACTAAAA CGTCAGCTGG CTTCAGAAGC CAGCTGACGT TTGTTTCTTC TACTGTTTAA	4260
CCAATTGGTG GCGTACCTTC TCGTTCCTCA GCAATCGCTT CGATTTGAAT CTGAGCACCT	4320
TTTGGTAAAT CCGTCACACC AATCACTCTT CTTGCAGGCG TGCCTTCAGG GAAGTACGTT	4380
TGGTAGACGT CATCTACTGC CGCCAGCTCT TCAATCTCTT TCACAAAAAT GTTCACTTTC	4440
ACCAGATCAG CTAAGCTGTG ATCCACACTT TCAATAATCG CTTAATATT TTCTAAACAT	4500
TGCGTCGTTT GTTCCTTGAT TCCGCCAGCG ACTAACGCAT TTGTTTTAGG GTCCAATGGT	4560
AGCTGAGCTG ATAGATGGTT GTAATGAGAA AAAGCAACAG TTTGTGTCGA AAATGGACAC	4620
TTCGGTGCGT GCTCTGTATT ATTTGCTTCA ATAATTAATC CATGTCGATC TTCAATCGCT	4680
TGTGGTGGCG TACCATCCCC ATGAGAAACC ACCGCTTCTA CTTGCACCGC CGCGCCCAT	4740
GGTAAATCAG CAACAGCAAT CACTGTTCGT GCTGGCATAT AATTGACCGT GCGCGCAATT	4800
CCTGAATCTG GGAAAAAGC GGCATGTACT TGATTGACAG CTTCCAGCTT AGATAAATCT	4860
TTTAAATAAA TGTTGATCTT CACAATATCA TCAAAAGGCA CATCAATACT GGTAAAAATG	4920

GCTTTTATGT	TTTTCAAACA	TTGCTTTGTT	TGCGTTTTAA	CACAGCCAGC	CACAACTCTG	4980
CCCGTTTGAG	GATCTATAGG	CAATTGTGCA	GACAGATTGT	TATAGTGAGA	AAACGCGACC	5040
GTTTGACTGG	CTAGAGCATC	AATTGGTGCC	TGAACAGTCT	GATTGTGTTAA	TTTAATTAAA	5100
TCCCCAGCTT	GAGGTGCATT	AGGGATAGTT	CCTTCCCCAT	TTGAAACAAG	TGCCTCCATT	5160
TGAACAGAAG	CgCcTAACGG	CAAATCTGCG	ACGGCGACCA	CTGTTTtagc	TGGATAATAA	5220
TGGGTAAAAA	AGGCCGCATA	GACCTCCTCA	ACTGCGGCAG	CATCCCGCAT	ATCTTTGACA	5280
AAAATGGTCA	TTCTGACAAT	GTCACTCATC	ACATGGTTGA	TACTTTCAAC	AACCGCTTGG	5340
ATATTTTTCA	AACATTGTGT	CGCTTGGGCC	CGAATGCCAC	CTTCTACCAA	TTGTCCAGAG	5400
GTTGGCTCAA	TAGGTAATTG	TGCCGACAAA	TGATTATAGT	GGGAAAATGC	GA CTGTTTGG	5460
GCTGCGAGAT	GATGCTTTGG	TGCACGCTCT	GTATTTCTTG	CTAAGACGAT	ATTTTCATTA	5520
CTCATAACTC	TACTCCCTTT	TTTAGCATTC	CAACAAAAAT	GATTCCGCTT	ACATTTTTAA	5580
GCATTAAACT	TCCAACAAGA	TCCTTTTTAT	TTTCCTGCTT	TTCTATTAAAT	GGTGATAGCT	5640
TTCCACCACC	ACTTCTTTCA	GTCTACAATA	AGAACAGGAA	CATTTCAATA	AGGCATTCCT	5700
AAATTGATTG	TTGGCCCGCA	AACAAATCTT	CTTCAGTTAA	AGAAGACTGG	CACGAATCAA	5760
GCTTGATTG	TGCCAGTCTT	TATAGaTAAC	TCGTTAGCTT	TAAC TAGTTT	TATttAaCCA	5820
CCATATACTC	ACGATAATTT	TCCGAGAACG	TATCCATTAG	CGTATAAAAT	CGTTCAAAAT	5880
CATAGTCACG	ACGAGGGTCA	TCATACCAAA	CACATTGAAT	ACTTCCAATA	ATTGGAAGAT	5940
TCGGGCTACC	TGCTGGCACA	TCTGTAAATG	CTTTTTTCGA	CATATTTGCT	AAAGCATCAT	6000
CATATTGATA	AATGCCAGAA	TCAACATTGC	CAGCGACCCA	ATACCACGCA	TCATTGGTAT	6060
TAAAAATTTT	GTGTCCTTTT	TGAACAAAGT	ACTCAGGTTT	GGCTACGTCA	TATCCGCTCC	6120
AACCGGCTGT	CCAATAAGAA	ATAATAATCT	CTGGGTCAAA	TGTACCGAAA	GAATCGTCGC	6180
TGTTGTAATA	AATGCCATCA	TTGAAGCTCA	TTGGCTGCAT	GCCCGCATCT	TTAATTATTT	6240
TAGCTAAATC	ATTCGCATAA	GCGACAAAAT	CTTTGTAGCG	CCCAGAAGAT	TGCAGTTTCG	6300
CCCAACCGCC	TGTATCGACA	TCATTtGCAT	ACTCATCGCC	GCCAAAATG	AAAATTTTAC	6360
TATGCGCGGA	AAAATAAGCC	ACGTACTTAC	TGATAATCGC	CTTTGTGAAA	TTCACTGCCT	6420
TTTGTTCC	TAAATCTACT	GTTCGTTTTG	AGCCATCAAA	AGCGGGATTT	TTAATCGCTA	6480
GTTTTTCCAT	GGCCACTAAC	AAGGCATCCA	TATGTCCTGG	ACTGTTAATC	ACGGGAATGA	6540
TGTTGATGTT	GCGCGCTTTC	GCAAAAGCCA	ACAACCGATC	CATTTCTTTT	TGTGTTAACG	6600
CGTTGCCATT	AGGATCATTG	TAATATGCGT	TGTTGCCTCG	TTGGATTGCT	TTTGAAACCC	6660
GGTTGTGGTT	GTATTTTTTA	CCATTCACAT	TGACCGACAT	ATCATCCAAG	ATAAACCGTA	6720
AGCCATCATT	TCCTAAAATT	AATTGGACAT	CTGTGTACCC	ATTTTGACTA	GCTTTGGCCA	6780
CTAATTCTTC	CAGTTGCTCC	ACGGAAAAAT	ATTTTCTTCC	CGCATCAATG	GAAAAGACAC	6840
TTTTTAAAGG	TTGTGTTTGT	TTCTCTTTTA	AATTATCCAA	AGCTGTTTGG	ACCTCCGTTT	6900

GCCAAGTAGT CACTTGTGCG GCTGTATACG TTTTCCCGC AAGTTCTGTG GCGACTGTCT	6960
CTTTTGTGTG TTGTAGTGCT TGAATCGTTT CTGGTGTATA GCTTTCTGCC TTTTCAATGG	7020
CTGTAAGTGC TTGATCCAAC GCACGATATA AAGTGGTTAA GTCAATGGCT AACAGGCCTT	7080
TTTCTAATTC AGCATCATAG GTGTTTGAAA TCCCTTGGCC ATAATCGCCA CCTAAAATAG	7140
CTTTATTAC ATCATAAAGC GTTGTGCTT GTGTCACTGC TTCTGCGGCT TTATTTTGTG	7200
CTTCAAGGTC TGTGTTGGCG TAGCTAATTC GTTTTAAATA GTGTTGCGCA GCcGCTTTCA	7260
TCTCTGCAAG AGAAACGCCT TTAAGTCTCTG CAATGGCCGT TTTTGTCCAA AGTAAATTAG	7320
AAGGTTTTAT CTGATTTAAA TCGTCTTCAT TATAGGTGCG ACCATCGCGA TCTAAGGCAT	7380
AGAGGAACAT GCCCCCTAAA TTATTTTCAT AAGAATAACG AGCTACTTTA TACATGTTAC	7440
TTCCATATA CGGCTCTTTT GTGTCAATCC AGCGTTGCG ATCTTGCTCT TCTGGGAAAG	7500
TTAATCCTGG AACAAATCGT TCCCCATTGA GAACCGGACT TAAATTATTT AATGCTCGTT	7560
GCGTGCGTTG GTCATCGCTG CCATACTGTT GATACGCGAG AAAATCGAAA CAGTCACTGA	7620
CATCTTGTA AGGTGGTAAA TATTCTGCAT TGGTATCATA CAAAATGGA CGATCCGTTT	7680
CCGACTTCGG TCCAATGTAT TTTGATAATG CACGAATGAC ACCATTAGAT AGAACAATAT	7740
CTTTTCACT TGGACGAGTT TCCATGTCAA TATCTAACC ATCAATTCCT AAATCATCGA	7800
CAAATTTGGT TAACAATCTT TTCGCATAAG CATCAAATTC TGCTTCTGTA GCGGTTGTTT	7860
CTGCATAAGG AACTTTTAAT AGCTCGCTGT AATCAATCCC ACGAACTAAA CGAACACCTC	7920
GTGCATGCAG TGcTGGCGCA TACTCATTTT TTAACGTATC ATAAAATGGC TGTGCAAGTG	7980
CTTCTTGTC TTTTGGCACA TAACTAAAGA CATTACAAAT GTCAATACCA TAAGGAATAT	8040
CGTGATCGT TAGCCAATT TCATCTGGCA ATGTTGTATT AACTCCTTGC ATCGTTTTGT	8100
CACGCCAAGC CCGATAATA ACCATAAAAT GTTTCTCTTG GGTGCGGCTCA TTGGCTTGGC	8160
TGATCATTGG TTGATGAAAC AGACTAAATA AGATAAACGC CAAGCCAACC AATAACGAAA	8220
GACCCAGCCA ACTGACAATT GTTTCTCTGT TTTCTCCTT TCTTGTCGAT AAATTGTTCC	8280
CCACTTTGAA CACCATTCCT TTCTGCACTC CGTTCATTTT TGTGTATCCA CCTAAAAGA	8340
AAGCGCATAC ACAACCATCA ATAGAATAAC AAAAAATAAC GAATTTGCAA AATAGAAACT	8400
GGAAAAATGA GTATTTTCTA TAAAAATCTT TTCCCGCCTA AAATCACTTT TGTTAGACGG	8460
CGTTTAAATG GTAAATTTTG AGACTTGCAA GAAAAACAGG CCTATTTTTC TTGCAAGTc	8520
CCTTTACTTA GTCGTGATCA ACCAATAAAT AATCACTGTT TGAATTTTTT GGGATCTACT	8580
CGTTTGAAAT TTATCTCATC ATCAATAACT GCTCTATGCT TTAAATcTAA ATGATATCGC	8640
ATATCATGAT TTATATTGA ATTTTTTTT AGATTATCAT TATATGTTT TAAAAATAAT	8700
TTTAAACTT TTTTCAAGTTG CCTTATACCT AATACTTTT TTGATATTT TAAACCATTT	8760
ACCTTTTTTA TAAACCATTT TTGATTGTAT TTCATTATAG ATGACAAATT AAAATAAGCT	8820
TGGAATACAG CTCCCGCATC TTTAATTTGA GTATTTTTTT TCAATTCATT GTATTCATTC	8880

AATAAATTCT TTAGCTCTTC ATAATTGTAA AAACCTGAACA TTCCATTCTT AACTTCTGTC	8940
ATCATTTTAT TCTGCCTAGT CCAAGATCCC ATGCATACAG AAACCTAATC AACTAAGGGA	9000
ATTTTGTTAT TCTCTATATG TAAATCTTTT AACTTTTGTG ATTCTTCATT TCCTAAAATT	9060
ACGTATTTAT TAATATAGTC TATTACCCGC CACTTACATG CTGTATTATT TATTTTAATA	9120
ATGTCATTCA TTTCTTTGTT TACGGAAACA TAGTATGGCA CGGGTAAGCC ATATTTTCTT	9180
GCAATACTAT ATCTATGTTG CCCATCGATG ATCTGCATTG TGGAGTTTAC TATAATGGGT	9240
CGAATAATTC CTTTTGTAG AATAGAATTT TCCAGTTTCT TGTTCAAGTGT AACTTCACGA	9300
TTACTTCTTA AAAAGCTGAA CATTTCATAT TCTTCCGTTA TATAAATATC ACTCATTTCT	9360
CTATCCATTT TTTATCTCCT TTATTAATAA CATACTTTGA ATATCTCTTA TTAAGGCTT	9420
CCCATCAATT TCGATTTCAT ATTTAATTGC TTTGTTTGAA TTTCTATGCA ATTTATAATT	9480
ATATGCTTTC AAAAATCTT CAAAACGAC TCCTTGGTGA AAAATTCCTT CTATAACATT	9540
TTTAGTTTTT ATTTTAAATC CATTAATTAA CCTATAATAA TCGAAATCTT TTACAATAAA	9600
TAATCTACA TAAGCAAAAA ATGTTTGCTG TGAACCTCTC AAGTTCAACT CTTACAAAA	9660
GTTTCCGTAA GAATCTAATA ATCTTATAAA TTCCTTGTAG TTATAAATG AAAATTGACC	9720
TGTTTTTACT AATTCTTGTG AATTACCATC ATAGTGTGTTA TTTCCCATAG CCAAAGATAT	9780
TAAAGAAGAG GGTGAGATAT TTTTATACTT CTTTACAATA AACTGTAGTT TTTTATATTC	9840
GTTCATACCA TCCATCGAAT ATTTCTGATG AAATCTCCAA TTTCCAGTCT TAATGTCGAA	9900
TTAAGCTCAA TTACATCATC TAAGTTCAGA CCTTTCGAAA TAATATAAGG AATGGACTTA	9960
TTCAATTTTC TGGCTATCAA AAAACGAGTT TGTCCATCAA TAATTTCAA TTTCTCATTT	10020
ACATCTATAG GAACTAAAT ACCTGATTTT TCTATAGATT CTATTAACT ACGCTTTTCT	10080
TTAATTGCTC TGTTTCCCTT TAAAAATTTA AATTtTGAAT AATCCTTAGT TTCATACAA	10140
TTACCAATCA ACTTTCCCAT ATTCTTCACT CCTTATAAAA TATATTTTAA ATTTATATAT	10200
TTTATAATAT CGTCTAAAAA AAAGAAATGC AATTCTTTTT TAATCAATAT AACACCAACC	10260
ACACATTAAA ATAACATTTT GGTTAACCAT CCATGTTTAT TAATTCGGCG ATTTTCACAA	10320
CAATTTTCCC TTGCGCTCGT CCTGTCAATG AATAATCTAC CGCTTCTTGA ATTTGGCTAA	10380
AAGGAACGAC TCGGTCAATT ATCGGCTGTA ATTTTCCCTG TTCAATAAAC TCAGTGAGTA	10440
AAGCTAATTG CTCGCCATCT GGACGCATAA ACAAAAAATG ATAACCTAACA TCTGTGGCTT	10500
GCTCCAATCG ATGAATCTTG CGGGTAGCTA TTTTAAAGGC CCATTGTTTC CAAAGCGCA	10560
AGCCATACTC TTTAGCAAAA CGTTCGTTGG GAATGCCTGA CAATGTAACA ACTTTTCCCT	10620
GAGGTTTAA CACTGAGAAA GCTTTTCTA AGATTGTCCC CCCATTGTA TCAAACACAT	10680
AATCATAGTC GGATAAACT TCTTCAAAAT TTTGTGTCCG ATAGTCAATC ACTTCATCTG	10740
CTCCCAACGC TTGAACCCAT TCTTTATTTT TACTACTCGT TGTGGTGGCA ACGTAAGCGC	10800
CTGCTAGTTT TGCTAATTGA ATCGCAATGG TTCCAATCCC TCCTGAACCT GCTTGAATCA	10860

GGACTTTCTG GCCTGGCTGT ACATTCATAA TATCATGTAG CGCTTGATAA CTTGTCAAAC	10920
CGACTAACGG AATGGCCGCA GCTTCTTCAA ACGTTAAATT TTTCGGCTTC ATAGCTACAG	10980
CCGCTTGATC CACTGCAATA TATTCAGCAA AAGTTCCAC GCGATTTTGG GGTACTCGCC	11040
CGTAAACTGC ATCGCCCAGC CTAAAATTTT GAACATTTT TCCAACAGAG ACAACAATCC	11100
CAGCAAAATC ACTGCCTAAA ATCAGTGGCA TTTGATAATC TAAAAGCATC TTTACTTTGC	11160
CATCTTTTGT TTTTAAATCA ATCGGGTTGA TACTGGCAGC GATAATTTTG ACCAAAACAT	11220
CATTGTCAATG AATGGTGGGT AAAGGCACCT CCTCGATTGC TAATTCtTTT TGTCCGTATT	11280
TATGAATTAA AGCGGCTTTC ATCATTGAAT CTTTCCTCTC TTTCTATCCC TTTTATCGGT	11340
TTAAAAGTTT AAACCTTTAA TCAATTAGTA TCTTAACGAA GAATGGGATC CACGTCAAGT	11400
TTCAGCCCTC AGCGTATTTT ATTGACAAAA AACACGAAAA ATGGTTTAAT CGTTTAAAAA	11460
TACGAACCAA AGAGGCGAAA GAATATGAAA GAATTAGAAG AAATTCAAAC AGAATTTACT	11520
GCGCTTAGTG ATTTTTTAAT CGCATTGGGT GATGAAAAAC GCCAAGCAAT TATTATTGCA	11580
CTGTTAGAAG ACCAAGCCTG TGTGGCCGT CGTGTCAGT AGTTAACAGA GGCGACACAG	11640
CTCTCACGCC CCGCTGTTTC TCACCACTTA AAAATTTTGA AACAGCCAA ATTAATTGAT	11700
TGTCGCAGTG AAGGAATAA AAATTATTAC TCACTTTCTC ACGATACCAC CAAAATTGAG	11760
CAACTCCAAC AACTATTGAC CCACATTACG TCGATTATGC CAGGGAGGAA GCAAGCATGA	11820
AAAAAATTCT CGTGGTCCTC ACAAACGTCT CCCGTTATCA CGGCACCGAA GAACCAACGG	11880
GCCTCTGGCT AGGTGAAGCC ACGGAGTTCG TGGAAGAAGT AACCAGCA GGATTTTCAG	11940
TTGACTATGT TAGTCCCCAA GGAGGCTATG TCCCTTTGGA TCCCCGGAGT ATGAAATATG	12000
TAGATTCATC CATTATGGCA GTTTATGAAA GTGCTGATT TCAAGAACGA GCGCTAGCAC	12060
ATTCCCTCTC TCTTGAAGAA ATTGAGCCAG ATGAATACGC AGCAATTTAC TTTACAGGAG	12120
GCCATGGCGT TATGTGGGAC TTTCTGATA ATCCTCAGTT GCAAGCAATC ACTCGTAGCC	12180
TTTACGAACA GGGCAAATAC CTAACTTCTG TTTGTCACGG AATTGCTGGC CTTTTAAATG	12240
TCAAAGATGC GTCAGGAAAT TATTTGATTG CAGATAAAAC CATTACTGGC TTTACTACTG	12300
CAGAAGAAAT TTTAGCCATG AAAAAAGCAG TTGTTCTTTT CTAAACGAA ACAATTGCCA	12360
AAGAACATGG TGCTACTTTT GCGAAAAAGC GTCCCTATAA AGAATTTGCG TATCAAGATG	12420
ACCGTATAAT CACTGGGCAA AATCCTTTTT CTGTCCGCGC TGTCGCAAGA TTGTTAATTC	12480
AACAACTATC AAAATGAAAC AACAGCGACT ACTTTTAGGT AGTCGCTGTT GTTTCATTGC	12540
TTCTTCTTTT TTCCGAAAGC CGAAGCCCTA ACATAAAAAA TTTTGTGATTA GCTAAAGAGT	12600
AACATCGCGC CAATTAAGAT AATTAGGTTA ATTAAGCCTA AGACAATCAT TGGTTTGGCC	12660
ATGAATTTTG TCCAAGTGGC ATATTGAACT TTGGAAATTT CCAAACCACC CATTACGACA	12720
CCAGATGTTG GCGTGATCAA ATTAATCAAG CCACAACCTG AAGCAAAAAT CATAACCATC	12780
ACTTCTGGAC TTAAACCAAT GTTTTTAGcT AACGCACCCA TTACAGGAAT AGAAACGTAC	12840

GCTAAACAG	AAGTTGAAGG	AATTAGGAAA	GATAAACCTA	AGTATAAAAC	AAAGGCACCG	12900
ATGACAAATA	AGACTGGGGA	GAGTCCTTGT	AATAAGCCTG	CCGCTTTATC	AAGGATAAAT	12960
GAATCTAAAT	ACGTTTTAGA	CATCAAAACA	GATGCCCCGC	GAGCAACTAC	GATAATTAAA	13020
ACAACGGATA	ACATATCTTT	GGCACCGTCG	ATAAATGCTT	CAATGGTTTC	TTTTTCAGTA	13080
AAGCGACCAA	TCAAGGCAGC	AATGATCCCT	AAAAGGAAGA	ACCACATCGC	TAAGTCGCCA	13140
AAGTACCAGT	CACCGAAACT	GTTGCCTGTT	AAAACAGACG	TCCAGCCTTT	GAAAATCGTC	13200
ACTCCAAAGT	CGTTCCATGG	AATTAAGGAA	ACAATCATAA	TGACAAAACA	GAAAGCGAAT	13260
ACCATTAAGA	TTTTTTTATG	ACGATTCGTA	AATGGCATTT	CTTTCGACGC	TGCTTGTCCA	13320
TACGTTTTTT	CCATATCCTC	TTGTTCTTGT	AAAGATAAAA	TTGTTGAACC	TTTGTGAGCC	13380
TTCACTTTTT	TCGCATAACG	CATCACAATG	AAAATAGAAT	AGCTAGTAGA	GGCCGCCCAC	13440
AAAATAGCAC	CGACAATTAA	AATAATTCCT	GTATTCGGCT	GAATGCCAAT	ACCTCGCAAC	13500
GCATCCATCG	CTACCCCTGT	GGAGAAAGGG	TTAACCGTTG	AACCGATAAC	ACCTGATCCT	13560
GCACCCAATA	GCACAGTACC	CACTGCGACA	AACGTATCAA	ACCCAGCGGC	GACCATGGTA	13620
GCACTCAATA	AACCATAAAA	AGGAATTGTT	TCTTCTGCCA	TTCCATAAGT	CGACCCACCA	13680
ATAGAAAATA	AAATCATTA	AATCGGAATG	ATTTTCAATT	CGTTCCCTTT	TAACTTTTTA	13740
ACGACTGATT	GAATCCCTGC	TTCAAGAGCC	CCAGTTGCG	TCACAATATT	TAAAAAGCCA	13800
CCGAGTACTA	AAATAAAAAT	ATTAATTTCA	ATGGCATCTT	TAAACCCATT	GAATGGAGCC	13860
ATGATAATAT	CCGATAATTT	GGCCCCTACC	ACGTGGTCAA	CCATTTTGCC	ATCAATTTCT	13920
CTTGGGGTAA	ATGGCTGGCC	ATTTAGCAAG	ACCGTTACGA	TACCTAAAAC	AATTAAAATA	13980
ATAAATAAAA	TACTAAAAGA	CGAAAGCTGT	TTTTTGTTTT	TCTTCTTTTT	TTCTGCCATA	14040
TCCTTTTATA	CAGGAACCTA	CTTTCCTGTA	TCTCCCTTCT	CTATGATTAA	TTACTCAATC	14100
GATCAATAAA	GAATGGCGTA	TCTAGAAAAG	TTAAGAATTT	ATGACTGTAG	CTAATACGGT	14160
TCATTTGGAC	TAATTTTTTC	ATAACACTGC	TTGCTGTTTC	CCGCGTAGTT	CCACTCATTC	14220
GTGCAATGTC	ATTAATGGTA	ATTGGACAAG	TAATTGCAAC	GCGTTCTCTG	TTGTCAATTC	14280
GTTCTCCTAA	ATCATTTAAT	AACATCGCCA	GTGATGTCAC	AATCCGTTCA	CATGCATCAT	14340
TGGTTGTTCC	TTTTTGGATT	TTGATCATAC	TTTCTTTTAA	AAGGACCGAT	TGTTCTTGAA	14400
TCCAATGAAT	CAATTGCTGG	GGATTATTTT	GGAGAAAACG	CTCAAACAGT	TTAACTGGCA	14460
CAAACACAAC	CTCAATATCC	GTATAGGCAA	TCGCGGAGTC	GTAATGAACC	TCATCTTGAA	14520
ACAGTCCGAT	TCTCGGAAAA	AGAGCCTGCT	TTTTCATAAA	ATGCAGATAG	GAAAAATTAC	14580
CACTGCGATC	ATTTTTTTCA	ATTTTAATGA	CACCTGAAAC	CAATAAAATT	AAATGATTTT	14640
TCGGATCACC	TGGATAAAAA	AGGACTTGGC	CTTTTTTATA	GGAGCGAACG	AAACTCTCGT	14700
TGCGCAATTC	TGTTAGTTCT	GATTCAGAAA	AGTATTGAAA	ATCTTGATTA	TTTTTAGTGA	14760
TTATTTTCTC	TATCATCATC	TTAATTCTCC	TAAATACTAT	ACTTTATTTT	GTCACAACCG	14820



TTCCTACGAT TTCATCGCCG CTCATACTTC CTAAGTTTTC CAAAGAAGTG ATAATTGCTT	14880
GTTTATTGGG TTGGCTTTCT ACAAACCTGAA TGGCGGCTTC AATTTTCGGT AACATACTAC	14940
CTGGTGCAAA ATGACCCGCT TGTGTGTATT CTTCCAACTC TGCTACAGTG ACATTGGTTA	15000
ATTGTTTTTC ATCTGGTTTG CCATAATTAA TACACACATA ATCCACGCCA GTTAAAATAA	15060
CTAGTGCATC CGCATCCACT AATTCTGCCA GTTTTTCAGA AGCAAAATCT TTATCGATTA	15120
CTGCTTCGAC ACCTTTTAAT TCTTGTCTTA CGACAGGGAT GCCGCCACCA CCACATGAAA	15180
TGGTAATTAT ATCATTTTTT ATTAAGGTAT TAATAGTCTC AGCCTCGTGG ATGTCAATTG	15240
GCTTAGGACT TGGAACGACT TTGCGCCAGC CACGTCCTGC ATCTTCTTTA AAAATAGCAC	15300
CTGCTTGCAT TGCTTCTTTG GCTTCAGCTT CTGTTAAAAA TGGACCGATC GGTTTTGTG	15360
GATTTTTGAA TGCCTCATCT GCTGGATCTA CGACCACCTG TGTTAAAACA GTAGCCACTT	15420
GTTTCTTGAT TCCTGCTTTG TTTAATTCTT GATTAAGCGC ATTGGACAAC CAGTAGCCGA	15480
TACTTCCTTG TGTCATAGCA ACACACGTAT CTAGCGGCAT CGCTGGATTT TTTTCAGAAT	15540
CAGCTGCTTG TTGTTGCAAT AATAGATTTC CAACTTGTGG ACCATTGCCA TGTGAAACAA	15600
TCAACCGATG CCCTTGTTA ATTAATGAA CTAAATAGGC AGATGTTTGA ACTAATGCTT	15660
GCTGTTGTGC ATGTGCGCTA GCATCATTTG AAAGGATGGC ATTTCCGCCT AGTGCAACGA	15720
CCATTTTTTT CCCCATTACA AATCCTTCTT TCCTATCATA ATAAAGAATA GGCCGCCCTT	15780
CAAAAGAAAC TCTTAAAGGG CGAAACAACCT ATTCTTTTTT CAATGTTTTT TCAGTTGCTT	15840
CCTACACACG AGGAATGAAT AAGTTGCCTA AAGTAGCTGC CATGATCGCT TTAATAGAAT	15900
GCATTCTGTT TTCTGCTTCT TCAAATTGAC GACCATATTT ACTGCGGAAG ACTTCATCCG	15960
TAATTTCCAT TTCAGTAATG CCAAATTTTT CAGCAACCAT TTCACCGTAT TGTGTTTTCG	16020
TATCATGAAA GGCTGGCAAA CAGTGTAAATA CGATTAAATC GCCATCCATG TTTTCGGTTT	16080
TTTCTAACAT AGCCATATTA ATTTGATAAG GTTTC AACAG ATTGACACGT TCTTCAAATT	16140
TATCTTCTTC ACCCATGGAC ACCCAAACGT CTGTATATAA CACGTTGGCT CCTTTAACGC	16200
CTTTAGCAAC GTCGTCTGTA ATCATTAAAT CTGCACCAGA TTCTTTCGCA AATTCTTTTG	16260
CATAATTAAC GACTTCATCA GAAGGGAACA ATTCTTTTGG TGCACAAATG CGAACATTTA	16320
CGCCTAAAAT AGCACCTGTC ACTAACAAGC TGTTTGCCAT GTTGTTACGG CCATCGCCGA	16380
CATATACTAA AGTGATTCCCT TCTAAACGAC CAAAGTTTTT TTTCACTGTC ATAAAGTCCG	16440
CAATCATTTG TGTTGGATGC CATTGATCCG TTAAACCATT CCAAACAGGA ACACCAGAAT	16500
ATTTGCTAA ATCTTCCACT ACTTCTTGAC TAAACCCACG GAACTCAATC CCGTCAAACA	16560
TACTACCTAA AACAATCGCT GTATCTTCAA CAGATTCTTT TTTACCTAAT TGAATATCGT	16620
TGGCGCCTAA ATACTCAGGA TGTGCCCCTA AATCAATCGC TGCTGTAGTA AATGCAGAGC	16680
GTGTCGGTGT CGAAGTTTTT TCAAATAAAA GAGCGATATT TTTTCTTCT AAATAGTGAT	16740
GTGGAATCCC TTTTTTCTTC AATTCTTTTA AGTGAATACT GAAGTCTACC AAATACTCTA	16800

ATTCTGCTCT TGAAAAATCT TTTTCTGCTA ATAAACTACG TCCTTGAAAT ACTGAATTCA	16860
TGATGATTCC TCCTATTTTT GGGTGTTTTT TTAAAGATCT TCACGGTAAA GTGGCATACT	16920
CATACAACGA GGGCCACCAC GGCCACGTAC TAGTTCACCT CCTGGAATGT AATCAATTT	16980
CACGCCTGCT TCTTCTAGCG CTTTATTCGT AATCGTATTC CGATCGTAAA CAACAACTTC	17040
CCCAGGGGCA ATTGCTAATG TATTTGAACC gTCGTTCCAT TGTTACGGG CTGCTGCTGT	17100
TAAATTTCCA GCGCCGCAAC GGATTAGTTG AACATTGTCT AAATGCAAGT ATTTGCATAA	17160
AATGTTATCT AATGTATCTT TTTCTTTTGT AATTTGGATG TCTCCATCTG CTTTTTCAGT	17220
GATCGAGTAA ACAACCAAAC CACCTTCGAT TTCTGGATGA ATCGTAAATT TATCATAGTC	17280
AATCATGGTA AAAACGGTAT CTAAATGCAT GAATTTACGG TGTTACCGA TATCAAAAGC	17340
CAAGATATTT TTAAACCCTA ATTTTTGTTC AAAAATATTT CTCGCAATTT TTTCAATTGA	17400
CGCGGCATCC GTTCTTTGAG AAATCCCAAT GGCCACTACT TCTTTTGAAA GAATTAATTC	17460
ATCGCCACCT TCAATTCTGG TTGATTCTGA ACGATCATAG ACTCTAGGAA CCTCTTTTCC	17520
AGCAAAACGA GGATGATAAT CAAAAATGTA TTGCCCAAAA ATGGTTTCCC GTTGTCGAGT	17580
TACTGAATAC ATATGATTTA AAGAAATCCC GTGCCCCATT GTCGCAAAGT TATCACGCGT	17640
GAAGTATAAG TTAGGCATTG GATCAATAAT GAATGGATAA TCCGCTTCAA CCATATCTGT	17700
TAAAAATTCT TGCTCATATT TAGGTAATTC TTGTTTTTGA ATGCCTGCAA TCGCTTTTTG	17760
AATCAGTTCT TCGTTGTCGT CAATTTCTAA CATTAACTCT CTAACTTTTT CTTTTGCTGA	17820
TTGCTGCGA ATATTGGCTT CTTCTAAGAA TTGGTCAATA AATTGTCGGC GGACCTCTTC	17880
ATTAATCAAC GCTTCAGCAG CTAAGTCCCTC TAAATAGACC ACTTCGATAT CTTTTGATCG	17940
TAACAACTCT GCAAATGCAT CATGTTCTGC TTGTGCTTTT TCTAAAAACG GAATATCATC	18000
AAACAACAGT CTCTCCAGAT AATCTGGCAT TAAATTTTCT AATTCCTTAC CTGGACGATG	18060
TAACATCACT GTTTTCAATT TCCCGATTTC AGAAAATACA TTAATTGGAT GACTCATGTG	18120
AAATAACCTC CTCAACTTAA TTACAAACAC AGAATAACAG ACCTGATTAA CCTAGGACGT	18180
GAGCCATTTG CCAAACAAAA TACCAAATAT ACAACGGTAT TCTTTCTTAT GCATATTTCT	18240
AGCCTAAAAG AATGCGCTTT CATAGATATT CTAAAAAAT AATGCATAAT TAAATAAAAA	18300
AAAATGCATT TATTCAAAAA AAATTATTTT TCCACACACT TCATAAAGTT ATAATCTTAT	18360
AGAGGKAAAA GGAAGTGAGC CAGAAGTCAA AACGACTTCT GGCTCACTTC CTTTTATGCG	18420
TTACTTTTTG ACAAAGGTGT TGGCTGAATA AATTGTTGAA AATACTGCTC TAGTTCTTGT	18480
GCTTTTTCTT CGGATGATGA AACAATAATA ACAGTGTCGA AGcCAGCGAC AGAACATTTT	18540
TTTTCACACA AAGACACTTC ATCAAGTACA GCTGAAAAcA ATGTTGCGTT GTCTGGTAAT	18600
GTATGAACCA CTAGTAAAAA GGCGACACGT TCAACTTTGG TCACTACCTC ACGAATTGCT	18660
TGTATCAACT GTTCTTTTTT TCGTTCTTCT AAAGAAGTGA ATTGTTCAAA AATCGTTAAC	18720
CGCATCTCGC CAGACTCATC TAATGCTTTG ACGACTTTTA ATTCCCATAA ATCCCGTGAA	18780

ATCGTTGCTT GTGCGACATT TATCTCGCGC TTTTTTAATT CTCTTAACAA CTCATTTTGA	18840
CTATGAATTA CGGTTGTTTG AATGATTCTT TTAATCAAAC TTTGGCGTGT TTGCTTATTC	18900
ACAAATAACA CCCTTCCTCC AGACATTTAC AAAGCAATAC TAGCAGAATC TCTCTAATTA	18960
GAATAGAAAA AGGGTCGTTT AGCTAAAAGG TTCTCATTGA TTTTGTCCGA TATTTGCGAT	19020
TTAAAAAGT GAGACTTGCA CAATTTAGTA AAACGTTGTA CAGTGATTTT aGGAATTAC	19080
TGGCACC GTT TTAATAACA GCAATGATGT TTTTTTGCAT GATTAAGTGA ACTGGAGGAA	19140
AGTAATGAAC CGATTAAAC GCCAAACGAT GATTAAGCGA TTGGTCACAG AAAATGAAAT	19200
CCATACACAA AGTGAATTAC TAGCATTATT GGAAAAAGAA AATATTACAA CCACACAAGC	19260
AACTATTCTT CGTGACATTC GTGAGCTGAA TTTAAACAAA GTAGACAATG GTCAAGGAGA	19320
AGCCTATTAT CGTATCTTGA ATAATTCTGC CCTTCGTCCA AAATATAAAA CAGATGAAGA	19380
ACGTCTTGTT GACATTATTA TTGAACTGG TGTTTCTTTA ACACAAATAG AGTTCACAAA	19440
TCTTTTAACC GTCTTACCTG GCAACGGGCA AGCCATTGGG GTTTTAATTG ATCGCATACG	19500
TATCCACAAC CCTCATATCT TGGGCTGCAT CGCTGGGGAT GATACCATCT TAATTCTTTC	19560
AAAAAATAAA GAAGATGCCC TTGAAGTTAA CAACTATTTC CAGCAGTATT TGTACCACCC	19620
ATAATTCATC ACTTAGTAAA ATCGACTATT TTTTGACATC ATtGCTATTT TTATTAAAGA	19680
TAAATAAAAA AAGGCGTCTC AGTGaAAGCT GAGGCGTTTT CTCATGAGTT TTTCTTTTGA	19740
ACACCGTCTT TTTTTTGTAT AATAATCAAA AAGCAAAGAA AgGGkTTCTT ATGACTGAAC	19800
AATTCATACT CACTTCAAAC GATCAACAAA CTCAGCTCAA TGTACGTCAT TGGCCCTGTC	19860
CTTCACCCAA AGCCGTAGTT CAATTGATTC ACGGCATGGC AGAGCATATC CAACGTTATG	19920
ATGAATTTGC TCGTTTTTTA AATCAATTAG GATTTACCGT AATTGGTCAT GATCATCTAG	19980
GTCATGGCGA ATCTGTAsaA CCAAACGCAC CGATCTATGG TTTTTTTGGA GAACAAGGGC	20040
CTGAAAATGT CGTTACAGAT ATTCATCAAG TGAAACAATG GGCTGTGAAT CGATATCCAC	20100
AATTGCCCTA TTTTATGATG GGTACAGCA TGGGGTCTTT CGCCTTGCGC AACTATCTGC	20160
AAGACTATCC TGTGACTGTG CAAGGAGTCA TTTTCATGGG CACTGGAACA AGTCCGTTAC	20220
CTTTAACCGC AGCATTGCCT TTTATTAAAA AAATGGCCGA GAAACAGCCG AAAAAACCAG	20280
CTCCGTTTCT TGATAAGCTG GCATTTGGTT CTTTATAGCAA AAAATTCCTT GAAGCAAGTT	20340
CCTTTAATTG GCTTTCTAAA AATCAAGCCA ATGTGGCTGA CTATGAAAAT GACCCATTAA	20400
TGGGCTTCGT ATTTACCAAC AATGGTTTTG CTACCCTTTT TTCGCTTGTT AAACGAGCCA	20460
ATCAAAGGAA TTGGTACCAA GCGATCCCAA AAGAATTACC TATTTTAATC ATTAGTGGCG	20520
CTGAAGACCC AGTTGGTGAT TTCAGCAAGG GGCCAGCGAA AATCAAAAAG CAATTAAAGC	20580
ATGCAGGTTT TCAGCACGTG ACGTTACGAC TATTTCCAC ATTGCGTCAT GAAATTTTAT	20640
TAGAAACGGA AAAAGCTACA GTTTTCAAG AAATTGGGCA TTGGTTAACG GATTTAACGA	20700
ATTAAAAAAG AAGGTTGATG TTGTGACATC AACCTTCTTT TTTAATCTAT TTTGTTAAA	20760

CCGCCCATAT AAGGACGTAG AACTTCTGGA ATCGTTACGG TGCCATCTTC ATTTTGATAA	20820
TTTTCCAAAA TGGCAGCCAC TGTTCGACCA ACAGCAAGTC CTGACCCATT AAGAGTATGG	20880
GCATATTGAA CATGACCCGT TTTATCACGA TAACGAATCA ACGCCCGACG TGCTTGAAG	20940
TCTTCACAGT TTGAACAAGA ACTAATTTCT CGATACGTTT TCTGTGCCGG AATCCAACT	21000
TCTAAATCAT AGGTTTTTCG TGCCGAGAAG CCCATGTCAC CTGTCGAAAG AGTAATCACT	21060
CGATAAGGGA GGCCCAATTT TTCTAAAATG TCTCCAGCAT TGTGGTCAT TTTTCTAAT	21120
TCTTCATAGG AGTGTTCTGC ATCACTAAAT TTAACCATTT CTACTTTATG AACTGGTGT	21180
AAACGAATTA ATCCTCGGGT ATCAGGACCT GCACTACCCG CTTCCGAGCG AAAAGAAGGA	21240
CTTAAAGCAG TAAAATAAAT CGGTAACCTT TGCGCATCTA AAATTCATT ATTATAGTAG	21300
TTAGTTAACG GGACTTCCGC TGTGGGGATC AGCGTCAAAT CACTTTCTGC AACTTGAAA	21360
ACATCTTCTT TGAACCTAGG AACTGACCG GTGCCAAACA TTGCTTTGCT ATTGACCAA	21420
TAAGGAGGAA TCACTTCGGT ATAACCATGT TCATTAACAT GTTGATCTAA CATAAAATTA	21480
TAGACCGCTC GTTCTAAACG AGCCCCTAAG CCTTTGTAAT AAAGAAAACG GCTACCAGAT	21540
ACTTTCGCGC CTCTTTCAA ATCTAAAATT CCTAAGGCGG GGATCCTCTA G	21591

## (2) INFORMATION FOR SEQ ID NO: 111:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6590 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

CAATTGCCCC GATTGGAATA AATGAAGGGC CTTGCGCAAC TGGTAATCTC ACACAAAAAT	60
GAGATTGAAT AATCGTGGCA AGTCCTGCTG CTATAAATGT TGCTTGGATT AAAGAAGACG	120
TTTGTGCTGT GCTAATCCCA ATAATTGAAG CAATGATCAC AGGTACGACG TACACGTCCA	180
TTGCTAAGAC ATGCTGAATC CCTAACATAG TGGCCTGCCC AAAAGATACT GAATCTTCTG	240
GCCCCACTGT TAAATGAAGA TTATTTTGTT TCTTTTCACT CATTTTTTAT ACTCCTTTAG	300
TTGCTGACAT TAAATTTTTA TCATGTACTT TTTCACCTTG TACCCAGACT TCTCGAATAT	360
TTTCTGGTCG TGCTAAATAT AAGATTTTTT GGAAAATATC AAGCAATGAT TCCCCTGGAT	420
TTTTAGGAAT TCGTGCTTGT GACAAGCTTG TATTAATAAT TTGAACATCC CAAGCATACC	480
CTTTTTCTAA TTTTCCAATT GGTAACCTTA GCGCTTGACC ACCACCAGCA GTAGCTAAAT	540
AAAACGCTTC ATTTACGGTA ATCCGTGCTG TTTTAACACC TCGTACTTCT GGTTTTTTTG	600
TAGCATCGAC GCCATCTTCT AACATTCTAG AAGAAATAAC CGCTTGCTTT AAATTGTCAA	660
AAAGACTTGG TGAAAAACCA CCTGAAAGAT CACTACCTAA ACCAACTTCT ACTTGATGTT	720
GATGTACTAA TTTTGCAATC GGTGTTACTG CATTACCAAA ATAAGCATTA GAGATTGGGC	780
AGTGCGCCAC AGCCGTCCCT GTTTCATGGA ATAAATTCAT ATCTGCTTCT TCTAAAAAGC	840

CTGCATGTGC CATCACTGcT TTTTCTGtCA ATAAACCAAA ATCATTCAAT GctTGAGCAT	900
CATTTTTTTC AAAACGTTCT TGaACAAATT GATGTTCCCA ATCACTCTCA CTACAATGAC	960
TTTGAACAkG GACTTGATAT TTCGCTGCTA ATTCTCCTAA TCCTTTAAst GCTTCCTCTs	1020
TACAACCTGG AaTGAAAcGC GGaGTTACAa CGGGrTACAC GCCTTGTTTT GTTTGCTGCG	1080
CTAATGTTTG AATTTCTTGA ATAAACCGTT CCGTTTCTTC TAATGCTTGT TGCGTTGTTT	1140
GATCACGATA GTATTCTGGA TTCTCATTAA GATCATCCAT TACAACCTTA CCAATTAAGC	1200
CACGCTGCCC TTTTTCGCA CAAATCTCTG CCAAACGACG GCTAGCTTCA TAATGAATGG	1260
TTGCAAAATA AAGGGCTGTT GTTGTCCCAT TAGCTAACAA TTGGGTACC AATTCATCAT	1320
AGACTTGTTG GGCAAATGCT AAATCTGCAA ACTTGACTC CAAAGGAAAC GTGTACGTAT	1380
TCAGCCATTC ATACAATGGA ATATCCAAGG CGGTCCCTGA TTGTGCCAC TGAGGTGCAT	1440
GGACATGTAA ATCAACAAAG CCGGTAAAA GATAAGAATC TTGGGGCAAT TCTTGAAAC	1500
GTTGTGCCTC TTTTGTTC TCTAGCACAG TTTCATACGC TTCAGCATCT GGTTCATTA	1560
TTTCCTGAAT TACTCCGGCC TCATCTATAC AAAAAAGTGT CTCTCATAA ACTTTAATCT	1620
CTTGCAAATT TTTACTAGAA AAACCTGTTC CTTTTATTAC AAAATGAAAC ATTTAATACT	1680
CCTCCTTACT ATCATATGAA GATTATACAT AAACTTTCGC TTAATAACTA TAAAAATGTG	1740
ATTTTTTTAA AAAAAATCCG TGTTAAAAACA AGTAAAAACC GAACATTCTA AGaATGrTTT	1800
ATTTTTTCTT TGGATAACA AGAAACTTT CTATTCTTAA GATTCTTCGG TATACTTGAA	1860
GAGAATCGTT AGAACTTTAG GAGGTCCTGA GATGATACAG ATCGaACAAA TTAGAACAGG	1920
TGTTATTCAA GAAACTGTT ATCTTGTTTA TAATGAAGAA GCGTTATTAA TTATTGATCC	1980
AGGTGCAGAA GGAGCAAAC TAATTCAAGA AATTACCGT TTAGGGAAGA AACCAGCCGC	2040
TATTTTATTA ACACATACTC ACTATGATCA TATTGGTGCG GTAGAAGAAT TGCGCCATCA	2100
TTATCAAATC CCTGTTTATG TTAGTCCATT AGAACAAGCG TGGCTTTCCA ATCCTATTTT	2160
GAATCTTTCA GGATTAGGCA GACATGATGA TATAGCAAAT ATTATTGTTT AACC CGCCGA	2220
ACATGAATTT AAAC TAACCG ATTATGAAAT TGGCGGCATG AAATTTTCCG TAGTCCCAAC	2280
ACCTGGTCAT TCAATTGGTA GCGTCAgcTT TGTCTTTGAT GATTTTGTG TTGTTGGGGA	2340
TGCATTGTTT AAAGGTAGTA TTGGTCGTAC CGATTACAC ACTGGTGATA TGCAACAATT	2400
ATTACACAGC ATTACGACCT ATCTATTAC TTTACCAGAA GAATTTCCCTG TTTATCCAGG	2460
CCATGGTGAT GCCACGACCA TTCAACATGA AAAAGCCACA AACCATTTT TTAATTAGAG	2520
ATTTATTAC TATTTAAAGA AACTTAGGAG GAGTAGCAAT GACTGAAGCA ACGACACTTT	2580
ATATTATTCTG TCACGGAAAA ACGATGTTTA ATACGATTGG CCGTACCCAA GGCTGGTCAG	2640
ATACCCCTT AACCAACAA GGAGAAGAAG GCATTTATCA TTTAGGCTTA GGTCTTAGGG	2700
ACATAGACTT TAAGGAAGCC TATAGTAGTG ATAGTGGTCG TGCCATGCAA ACCGCTCAGA	2760
TTATTTTACA AGAACATCAA AACCACCAA AGATTCTTA CCTAACCGAT AAACGGATTC	2820

GTGAGTGGTG TTTTGGTTCT TTAGATGGTG GCTATGATGG TGAAC TTTGG GGTGTGGTGC	2880
CACGCATTTT AGCTTTTAAA AGTTATGAGG ACATGATGAC CACGAAGATT ACCTATCGTG	2940
AGCTAGCAAA TGCTATTATT GAAGCCGATA CGGCTGACTG GGCGGAGCnT TATGAAGTTA	3000
TCCGTGATCG TGTTTGGTCT GGTTCGAAG ATATCGCTCA TCATCGAGAA AAAAATGGTG	3060
GCGGCAAAGT CATGGTGGTT TCCCATGGAC TAACAATTtC CTTnTTACTT tCCTTAATgA	3120
TGCAAtTTGC CTATGCAGAT GGcGCTGGAG AATGGTAGCG TGACGACCTT AACTTACGAA	3180
AAAGGAACAT TTACCATnCA AGGAATTAAT GATATTTCTT ATATTGAAAA GGGAAAAAAG	3240
ATCGCTGAAA AGCgCCGTTT GTTGTA AAAA AAGAGAAGGT TATCTGTCAA AATGACAGAT	3300
AACCTTCTCT TTTTATTTAT TCTTTTAAAC GTTCTACGTT TTTTCTTGAA ATTTGAATTT	3360
TTTCTAAGGT TTCCTTATAG TTTTTCGTG CATAAGCAAA AAATAAAACA TCGACTACAT	3420
ACAATTGTGC CACCAAGGAA ACAGTAGCCG CACTTCTCAA TGGCACGTCT TCCCCACTAG	3480
AGGTTTGGAG GACGATATCA CTTTTTTCAC CAATCACTGA TTTTTCATCT TGTGTGATAC	3540
TAACAATAGG AATCCCTTGc TGctTCGCTA AATCAGCTAA AGTACTTGCT TCTTTATTCG	3600
TTCCAGAATT AGAAATAACA ATAAAAACAC TAGGCTTTTC TGTTGAGCCT AACATGGAAG	3660
CAAATAAATG ATGATCCAAT GTTGTAATA CCGTTCGGCC TAATCGCGTA AATTTTTGAT	3720
AAATATCTTG AGCAACCAAC GAAGAGGCGC CCAACCCATA GACAAAAATT ACTTCTGTTT	3780
TTTCCAACAA ATCAACGGCT TCATTCACTG CGACATCCTC TAATGTCTGc aCgctTCGCT	3840
CAACCACATG AACAAAGCGC GCCTGTA ACT TTTGTTAAT GTCAGCTGTT GATTCCCCTT	3900
TAGTAACTTC AGTATACATT TGCTGCTTCA TTTGGCCAAG ATTTGCAGAT AATAAAAGTT	3960
TCAACTCTGT GAAGCCATTC ACATCCaTAG AATGACAAAA GCGAATAATT GCCGCTGGGC	4020
TTGATCCTGC TTGTTTCGCT AGCTCTTGCG CATTCA TTGA AATAACTGTT TCTGTATTTT	4080
TTAAATATA TTCTGCGATT TTTCGTTCTG ATTTTGGTAA TTGTTGTAGA CGATCTTGGA	4140
TAGATAAGAC AATATTTTGT TGCATAAACA ACCCCTCCTG ATACACCATT TCCTACGCAA	4200
TTAAATTTTA ACAAAGAAAA AGGGAGAAGC CAAGTTCTCC TTTTCTATT TATTATTTAG	4260
TTGTTTGCAT ATCTTTAGGA ATCCCAAAGA ATAACGTTGC TACAAAACCA CCTGCGTAAG	4320
CAGCCAATAA ACCTAATACA TATGCAAGCC ATTGGTTGTT GGCAATTAAw GGAATTAAAG	4380
CCACCCCACT TGGTCCAATA GCAATCGCAC CAACATTACC GAAgGCCCCA ATCACC GCAC	4440
CACCAATACC ACCGCCGATA CAAGCAGTAA TGAATGGACG TCCTAATGGT AAAGTAACTC	4500
CGTAGATTAA TGGTTCACCA ATACCTAAAA TCCCAACTGG CAAGGCACCT TTAATCATTT	4560
CAACTAATTT TTTATCTGAT TTACAACGAA TCCATAATGC TAAAGCTGCA CCGACTTGTC	4620
CCGCACCAGC CATGGCCAGA ATTGGTAATA ACAAAGTCAT CCCTGTTTGA TTAATCATTT	4680
CAATGTGGAT TGGCGTCAAA ATTTGATGTA AACCAAACAT TACCATTGGC AAsAAGGTTA	4740
AACCAAGAGT AAAGCCAGCA ACCATGCCAC CTTTTTCTAA GACAACATTG ATAATACCAA	4800

CTAAGCCATT AGAAATAGCT CCGGCTACAG GCATGATTAA GAAAATAGTT GCTAAGCCAA	4860
TCACTAAAAG TGAAATGGTT GGC GTTACGA TAATATCAAT TGATTCTGGG ACAATTTTAT	4920
GCAATTGTTT TTCTAACAAT GATAATAACC ATACGGCAAA AATAACACCG ATAATTCCGC	4980
CTTGTCTGTC AGATAATGTT CCCCCAGTGA AAATATTACT GATTGGAGCA TCCGGATTCA	5040
TTCTGTGTAA CATTGTGACT GCCCAATAA CACCACCCAA CGCTGGCGTT GCACCAAAAA	5100
CGCTGGCACT GTTAATCCCA GTGTAAAGAG CTAAATAAGC AAAAATCCCA TTTTAAATGA	5160
TATTTAAGAC ATCAATATAT TGTGCCATG AAGCTGAAAT ATCACCAGCA ACAACTAAGT	5220
TAGACATAAC AGCTGCAATC CCGCCAATAA TACCAGCCCC AACAAAGGCA GGAATCATCG	5280
GAACAAAGAT ACTAGAAATG GCTTTTAACA CTTTTTTAA TGGTGAAGTG TTATTTTGTT	5340
TCGCTTTTGT TTGTGCTTTC ATTTGAGCGG CTTTTTCTTC TACTAAATCT TTTCCAGATT	5400
TGCCTGCTGA AGCATCAGTT GTGGTGCCAT GATTAAATGG TTCTCCCAAT TTGACACCTG	5460
CTTGATCTAC CATTTCTTGT GCAACCTTGT TAACAGTCCC TGGACCAACA ACTACTTGTA	5520
ACGTATCATC TTCTACGACG CCCATAACGC CATCAATTTT TTCAATCCT TCAATATCCA	5580
CTTTGTCATA ATCGCGGATA TCCATACGAA CACGAGTCAT ACAATGAATG ACTTTGTCCA	5640
CGTTTTCTTG ACCACCAACT TCTTTATAAA TCCCATCAGC AATGCGTTTA ATTCTTTCTT	5700
CTGCCATGTT AAAACACCCT CCTAATTCCT TAATTTAATG TATCTTGAT AAATGATTC	5760
CCGTCAATTA ATTTTGTGT TGCTTCTTCT TTAGTACTAT CAGTTAAAT CATTACAATT	5820
GCTAACTTCA CATTTTCTTC TGCTTCTACA AATTTTAGTT CGGCAACTTC ATA ACTACAA	5880
CCCGTTGCTT CCATAATAAT CCGTTTGGA CGCTCCACTA ATTTTTCATT GGTGGTTTT	5940
ACATCAACCA TTAAGTTATT ATAGACTTTC CCAATACCA TCATGGAAAT AGTTGATAAC	6000
ATATTTAAAA TTAATTTTGT AGCTGTCCCA GATTTCAAGC GAGTCGAACC TGTAAAAAT	6060
TCAGGTCCTG CATCTACTTC AATCGGCATT TGCGCATATT TACTAATTTT GGCATTTTAA	6120
TTACATGCCA CAGTTGCGGT AGCTGCTCCC ACCGTTGTAG CATATTCTAA ACCGCCAATT	6180
ACATAAGGCG TCCGACCACT GGCTGCAATC CCAACAACAA TGTCATTTGC TGATAATTTT	6240
AGATCAACTA AATCTTGACG ACCTAGTTCT TTTGAATCTT CGGCACCCTC TACTGCGACA	6300
GTCATTGCTT TTTGACCACC AGCGATGAGG CCTTGAACCA TCTCTGGCTC GACGCCAAAA	6360
GTTGGTACGC ATTCGGCTGC ATCTAAGACA CCTAACCGAC CACTCGTACC AGCGCCCATG	6420
TAGATCAAGC GGCCGCCCTG AwTGAAGCyT TTGaTGaKGG CTTCTACCAC TGGTTCAATG	6480
GCTGCTAATT CTTGACCCAC TGCCaTTGCA ACTTTTTGGT CTTCTTGATT CATTTTTTGT	6540
AAAGCTTCTT TGACACTCAT CTCATCTAAG CCCATTGTaT TTTCATTTCT	6590

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3660 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

GACATCTGTC AAAACGACCG ATGTCCTACT CTCTAACGAT TAAGATAGTC CGCTAAATGC	60
GTTTGTTTTA GTCTTTTATA AAAGTTAAAT ACTGAGAATT TATTGTGTAA TTCGTGTACC	120
ACTTCTCCTT TTGATTGCTT CCGAAGCTAA TTCTAAAAGC CCGATAATCG CCTGGCGGTT	180
TGGTTTACTT TCTGCAAAAC TGATGGCCGC TTCCACCTTA GGTAACATAC TACCTGGCGC	240
AAATTGATTT TCTTCTATAT ATTGTTGCAT TGCCGTAACC GTAACATCGG TCAAAGCTTG	300
TTGATTGCTT TGACCAAAGT TAACAAAAC ATGTTCCACT GCAGTTAATA CCACAAAAC	360
ATCTGCTTCA ATGAGTTCAG CCATTTTTCG TGCCGCAAAG TCTTTATCAA TGAATGCATC	420
TACACCATGA TAGACTGGaC CATCATAGGc AaTAGGrACC CCGCCACCAC CACAAGCAAT	480
TACGATTGTG CCTGgCATCA ACAAGGTTC GCAGCCTAAC TTTTTCGTAA ATATCAATGG	540
TTTTGGnTGA AGGAACtACC CGTCGcCAGC CTCGaCCCGc ATCTTCyTGa TAGACGTTCC	600
CAGTTTCCGc CATCmACTGt TTGGckGTTT CyTCGtCATA ATAACCACCA ATTGGTTTTG	660
TTGGTTTTGT AAaCGCTGGG tCTTTTGGaT CCcACCACGA CTTGTGtAAT TAACGTAACA	720
ACTGGTTGTT CCCCTAAATT TCTGGCTACT AGTTCTTCAT CAATTGCTTG TTGTAAATGA	780
TAGCCAATAT AGCCTTGGCt CATtGCTGTG CATTCaGGAA ACGGCACGAC ATCTTCTGGA	840
GATTGACTCG TTTCTCAAA AGCTAAGCGA ATTTTGCCAA CTTGTGGCCC GTTGCCATGA	900
GCAATCACCA CTTGATGGCC TTGTTCaATT AAATCAACAA TGGATTTTGC AGCGGTTTGT	960
GCTTTTGCCA ATTGTCTTT CGAAGAATTT CCTAACGCAT TTCCACCTAA TGCAATCACT	1020
ATTTTGTACA TCTGTTTCCC TCCTAGCGCT TTGGTTCTTG TTGTGTGATA CAGTGAATAT	1080
TGCCGCCCCC ATAACAACt TCTACCGTAT TAACGCCGAC AATTTTTTTA TCTGGGAACA	1140
TCGTTTGAAC TTGCTCTAAA GCCAAATGAT CATTTTCATC ACCATACTGT GGCACAATCA	1200
CGCCATCATT TGTAAPTAAA AAGTTCATAT AGGATGCGAT ACAAATATCG CCGTCTTCTC	1260
GTGGCATGGT TCCTTCCACG AAATCAATTT TGAAGGAGCC TTTAATGGTT ACATTTTTGA	1320
CAGGACAGCA TAATTGTGG ACTTTTAATT GGCGTCCTTT AGCATCAGTC ATTTTAAGCA	1380
AACGTTGATA CGCATCTTGT GCCGCTTCAT AAAATGGGCT ATTTTGATCT TCAGTGTAAG	1440
TACAAGCCAC TTCACCAGGT GCAATAAAAC AArCTACATC ATCCACATGA CCATTGGTTT	1500
CTTCTGGGTC AATTCCATCA CCTAGCCAAr GAaCTTTTTT AACATTTAAA TAGTCGCATA	1560
ATTTTtGTTT GATTGCTTCT TTTGAAAGTT GTGGATTTCG ACCTTCACTT AATAAGCACA	1620
TTTCTGTAGT CAAGACTGTT CCTTGTCCAT CTACGTGGAA TGAACCGCCT TCTAAAACAA	1680
AATCATCTGT ACGATAAGAA TCTACATGCT CAATTTTACA AATTTTTTGT GCAACCAAAT	1740
CGTCTTGGTC CCAAGGGAAA TATAAACCAT CTACTAATCC CCCCCAAGCA TTGAATGTCC	1800



AATCCACACC	ACGGATTTCT	CCATGGTCGT	TAATGACAAA	TGATGGTCCA	CAATCACGAA	1860
CCCAGGCATC	ATTGTTTGAC	ATTTTCGTAAA	CCGTAATTTC	TGGCGGCAAT	TGGCGACGGC	1920
AATTTTGAAA	TTGTGTGTGA	GAAACCACCA	CATTCATCGG	TGTAAATTGA	CTAATTGCTT	1980
TGGCCACATT	AGTAAAAGCT	TCTTGGACAG	GTTTACCACC	ATCACGCCAA	TTATCCGGTC	2040
TTTCTGGCCA	AATCATCCAA	ACTTTCCTCT	GTGGTTCAAA	TTCTCCTGGC	ATTCTAAAGC	2100
CATCTTGtTT	AGGGGTACTT	CCTACAATTC	GTTTAGCCAT	TTTCTTTCGT	TCCCTTCTTA	2160
ATTTTTTTCA	CGTAAATAAT	TACTTCACCA	AACAGAATAA	AAATGATTGC	CCCAATTGTG	2220
ATTGGTAATT	TTTCATTTAA	TGCTGCGGGA	CTAAAATCTA	AAGGcACTGC	TGTGAAAATC	2280
AGTGAAATAA	TAATCATAAT	CATTGGCAAA	ACCACTAAAA	CTTTTAAAAA	TCCTGGTTTT	2340
CCGTAACTT	TAAACGGACG	TGGTGTATCG	GGATCAATTT	TTCTAAGCTT	ATAAAAAGCT	2400
GGGAAAAC TG	GGATATACGA	TAACAAGAAT	AGGACGACAT	TCAAGGAGAA	AAAGGCCAG	2460
AATAAATCTT	GATTAGGTAA	AATTGGTGCT	AATACAATGA	CCACCGATGC	CACAATGCCA	2520
TTCATCAAAG	CAGCCCCAAT	TGGCATGTCA	TTTTTCTTAC	TACGTTTTTC	AAAAAATTTA	2580
GGCATGTCAC	CATTTTCAGC	AGCATAACAA	GCTGTATTAT	TGACTCCTAA	TGACCAAGAA	2640
ATCATGTTTC	CGAACACCGT	AAGAAGGAAT	AGGAAAGCCA	TCAACATAAT	GAACCAACCT	2700
TCTGTGAGC	CTGTCAATAG	TTTAAAGCTA	TCCATCATAC	CACTGCCGGT	ACTAATTTGA	2760
TCGGTGGGAA	TCGCTACCCC	AATTCCAAAG	GCTGAAAAAA	TATAAATGGC	CGCAATGACT	2820
AACCCCGCTG	CTACAATAGC	TTGTGGAATT	TGCTTCTTAG	GATTTTCCAT	ATCATCTGCA	2880
AACGTACAGA	TGACTTCAAA	CCCTAACAAA	TTAAAAATGA	TTACAGAAAT	AAAAGATAAA	2940
CTCCTTAAGT	CAAACGTTGG	CAGCATTGAT	TTCAAGGTAA	ATTCATTGGC	AACACCTTTC	3000
GTCAAGGCTA	CATACAATCC	TAGCCCACCA	ATTAAGACAG	CTAAAACCAT	CTTGATTACG	3060
GCAGCTCCAT	TTAGAATCCA	GATACTGTCT	CTGACTGGAT	AAAACTAAT	CCAAACCACA	3120
ATCCAAGTAA	AAATTAATTC	AACAATAATC	ATCATCGGTG	TTGAAAACCT	AATACCTGTA	3180
ATTGTGCTTA	ATAGTTCTGG	CGTCATAACG	GCTAATGAAG	CTAGCCACAA	TGGATAATTA	3240
ATCCAATAAT	ACCAAGAAAC	TCTGGCACCC	CAACGATGAC	CATATGCTTT	GGTCACCCAA	3300
TCATAGATGC	CCCTTCACC	AATATAAGTT	GTTCCTAATT	CTGATGAAAT	TAACCCATAT	3360
GGCAATAGGA	ATGCAATTAA	TAAGAAAATC	CACCAAAAAA	ACTGAGAATT	ACCAATTGCA	3420
GCAACAGGCG	CCGTGCTTC	AGCAACGAAA	ACCACACAGA	TCACTGATAA	AATGGCACTC	3480
GTAAACTAA	ATTTCTTTTT	CCCCTCCATG	AAGACTCTTC	CTCTCTTTAT	AAACTGTCCA	3540
CTGATTCATT	TTTTCAATCA	GTGGACAGTT	GGCCGTTAAA	TGCTTTGAGT	GTCTAAGAAA	3600
ACTTCTAGTT	CTTTTTTCGC	TTCTGcTTGC	TTGATTAAAT	CATAAGGATT	TTTgCTTCAT	3660

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 827 base pairs

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

AATGACAACA TTTACAAAA AGTTTGAGGC AAAAATTAAA ACGGCGGACA wTAAaCTTaA	60
AGTAAAAaTT TTGTTTTTAA AAaTTGCGAA CTTAATAAGT GGGCTTTTTTC TTTTTCGTTC	120
AACCATTATA AATAATCCTA ATGCGATTAC TGCAATAATG AACaATAAaT AGGATTGTAA	180
CGAGCCaAAC CCGCTCTCTT GTCCGAGGAA AATCCC GCCG AAGAACGTCA TAATCGCAAT	240
GGCAATACAA GCAAATCCTG AAAAATCGAT TTTTCTTTT GTTTTTGTAA TATCTTTTGG	300
TAGGAAC TTT TCGCCAATTA AAATTGTAAC TAGCCCCACG GGAACATTAA TCCAAAAAAT	360
ATACGACCAA GAAAAGTTTG ATAAGATCAA TCCACCAATC CCAGGACCAG CAATCGCGCC	420
TAAAGAGACA AAGGCACCTA CTGCACCAAG CGCCCGCCCT CTTTCATTTA AAGGGAATAC	480
TTCCGTAATA ATTCCTGAAT TGGTCGCCAT GGTCATGCTA GCACCAATAC TTTGAACGAC	540
TCGCGCAAAT AATAGAAAGC TCAGCGATTG ATTAAGCCA CAAAGTAACG AACCAATCGT	600
AAAAATAACC GTGCCAATTC GATAAACTTT TATTTTCCA AACTGTTCG CAATTTTTC	660
AAATAATAAT AAACAAGCAC AGACAACCAT TAAATAATA GATACGACCC ATTCCGCTTG	720
ATTCATGGGT ACGGACATTT CTTtTGAAAT AGTCGGTAAa GCAATATTTA CAATGCTGGA	780
ATCTAATGTA GACATGAAGG TAAACATAGC CACAGAACT AAGATCC	827

## (2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13574 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

AAGGCATGAA GTAATnGGGA TGGATTTATT AATCGrATGA AGyCatATCa GGAAGAAGAA	60
GCAGTTTGCT TTACAGATCG TTTGAAAGAG CAGGTGGAGA AAAATGGCTA GAAAAAGTCG	120
TATAAAAACA CCGATTGcGG AACAAGCACC TACTATCAGA AAGAATAATT TTGAAGAGGT	180
TTGCTTAGGC TATACATTAG CAGAAGGGCA AGAAGAAGCG ATTCGCTGTT TACAATGTAA	240
AAATGCACCT TGTATTGCTA AATGCCCCGT GATGATTGAT ATTCCTGGAT TTATTTTAGC	300
TATTCGTGAA GGAAACATGG AACAAGCAGC CGATATTTTA AGTAAATATA CGAATTTACC	360
TGCTATTTGT GGACGCGTTT GTCCGCAAGA GAAACAATGT GAACAAGTTT GTAAATTAGG	420
GAAAGCCAAA AACTTTGAAC CAGTGGCAAT TGGCAAATTA GAACGTTTAG TTGCCGATTG	480
GGCATTAGAA AATCACAAGT TTCCTAAGAA AACAGCGGAA CCAACAAAAG GGAAAATCGC	540
AGTGGTCGGT TCTGGTCCTT CTGGGTTAAC AGTTGCGGGG GATTGTCAA AGCTAGGCTA	600

CGAAGTAATT	ATTTTGAAG	CGCTTCATGA	AGCTGGTGGC	GTATTAACCT	ATGGCATACC	660
TGAATTTTGA	TTACCGAAAA	AAATTGTAA	ACAAGAAATT	GCGAGTATTG	AAGCCTTAGG	720
TGTAACAATA	GAAACCAATG	TGGTAGTAGG	TAAAACGATT	ACGATGGAAG	AAATAATGTC	780
AGAGTTTGAT	GCCTGTTATC	TATCGGTGG	TGCAGGGGCA	CCAAATTTTA	TGGGAATCCC	840
AGGAACTAGT	TTAAATGGGG	TTTATCTTC	TAGTGAATTT	TTAACACGAA	TCAATTTAAT	900
GCATAGCTAT	GAGTTTCCAG	AATATGATAC	ACCGATTAAA	CGTGCTAAAA	ATGTTGTTGT	960
GATTGGCGGC	GGGAATGTTG	CAATGGATGC	CGCTCGTTCA	GCCAAACGAT	TAGGCGCTGA	1020
AAATGTGAAC	ATTGTGTATC	GTCGTTCTG	AGAAGAATTG	CCAGCAAGAA	TTGAAGAATA	1080
TCACCATTCC	CTTGAAGAAG	GAATTAATA	TTATTGGTTA	ACGAATCCAA	TTGCTTATTT	1140
AGATGACCAA	CAAGGAAATC	TAGCCGGTGT	TGAATGTGTC	AAAATGGTTT	TAGGTGAACC	1200
AGATGCGTCT	GGAAGACGAC	GGCCAGAACC	AATTCCAAAC	AGCACGTTTA	CCATTCCCTGC	1260
TGATGCAGTA	ATCGAGGCAA	TTGGCCAAGG	GGCAAATCGA	GTCTTGTTAT	CAACATATCC	1320
AGAAATAGAA	CTAAACCAAT	GGGGTTACAT	TCAAGCTGAT	CCTAAACAG	GAGCAACCTC	1380
AATACCTGGC	GTTTTGCTG	GGGGCGATAT	AGTCACAGGT	GCTGCCACCG	TTATTTTAGC	1440
GATGGGCGCT	GGGAAAATCG	CAGCTAATGC	AATTGATCAG	TATGTCAAAA	CACAAAAAAC	1500
aACAGTCACA	AGCAATGTTT	AGGAAAGATA	AAAGGTGAAA	AAATGAAAAA	AACAATGGAT	1560
GGAAACACCG	CAGCGGCTTA	TATTTATAT	GCGTTTACAG	AAGTTGCCGC	TATTTATCCT	1620
ATCACACCAA	GTTCAACAAT	GGCTGAGTTG	GTAGATGAAT	GGGCGGAAAG	TGGTTTAAAA	1680
AATATTTATG	GTCAAAAAGT	TCAAGTTATT	GAAATGCAAT	CAGAAGCTGG	GGCAGCGGGA	1740
GTAGTTCATG	GTTTCTTAAA	AACAGGTGCG	TTAACCATA	CTTATACTGC	TTACAAGGT	1800
TTACTGTTGA	TGATTCCCAA	TATGTATAAA	ATCGCAGGAG	AATTGCTGCC	TTCTGTGTTT	1860
CACGTGGCGG	CTCGGGCAGT	AACGACAAGT	GCGTTGAGTA	TTTTTGGTGA	TCATGGCGAT	1920
GtATGGCTA	CCCGCCAAAC	AGGTTTTTGT	ATGTTAGCAG	AATCGAGTGT	TCAAGAAGTG	1980
ATGGATTGTG	CGGCAGTTGC	TCATTTAGCT	AGCTTAGAAG	GTAGTTTACC	TTTTGTAAAC	2040
TTTTTTGATG	GTTTTAGAAC	GAGCCATGAA	CTACAAAAAA	TCAAGTGAT	TGATTATGAT	2100
GACCTGAAAG	AAATGGTGAA	TCAGGAAGCT	GTAACCTGTT	TTAGAATGGG	AGGCATGAAT	2160
CCTAATCATC	CCACTGTTTC	AGGGACCGCT	CAAAATCCTG	ACATTCATTT	TCAACAAAGA	2220
GAAACAGTCA	ATAAAAATTA	CGAAGAAATG	CCAAAAATTG	TTCAAAAATA	CATGAAAAAA	2280
ATTAACAATC	TTCGTGGAAC	AACGTATGAT	TTAACCGATT	ATTACGGGGT	TGAAGATGCA	2340
ACAGAAGTCA	TTATTTCAAT	GGGGtCGgct	TCGCCAGTAA	TCAAACAAAC	AGTCGACTAT	2400
CTCAATCAAC	AAGGACGCAA	AGTTGGTTTT	ATCAATATTC	ATTTATATCG	TCCATTTTCT	2460
ACCGAAAACC	TTCTTGAAAA	ATTACCAGCT	ACAGTAGAAA	AGGTTGCGGT	TTTAGATCGA	2520
ACAAAAGAAT	CGGGAGCCGA	AGGGGAACCG	TTATTGCTAG	ATGTTCAAAG	CGCTCTCTAT	2580

CAACATGAAA	ATCGACCAAT	TGTGATTGGT	GGACGTTATG	GGTTAGGCTC	TAAAGACGTA	2640
GCGCCTAATC	AAATTAAAGC	GATCTATGAT	CATTTGTTAT	TGCCATTTAA	AGAATTAAAA	2700
CAACGTTTTA	CGGTAGGGAT	CGTTGATGAT	GTAACCTATC	AATCATTGCC	AGAAGAACCT	2760
ACTTTAGATT	TAACGCCAGA	TACTACCTTC	CAAGCAAAAT	TTTGGGGCTT	TGGTTCTGAC	2820
GGAAGTGTGG	GAGCAAACAA	GCAAGCGATT	AAAATTATCG	GAGATCATAC	CGACCTTTAT	2880
GCACAAGCTT	ACTTTAGTTA	TGATTCTAAA	AAATCAGGCG	GTTTAACCGT	TTCGCATCTG	2940
CGTTTTGGTA	AAGAGCCAAT	TACTTCCACG	TATTTAATTG	AACAAGCAGA	TTTCATCGCT	3000
TGTCATAATG	CTTCCTATTT	GCATAAATAC	GACCTGTTAA	AAGGTTTAAA	AGATGGCGGA	3060
ACGTTTTTAC	TGAATACTAT	TTGCGATCAA	GAAAAGATTC	ATCGTTTATT	ACCGGCTAAG	3120
TTAAAAAAT	ATATTGGGGA	GCACCAAATC	AAATTTTATA	TTATCAATGC	GGTTGAGTTA	3180
GCTCGCAAAG	TAGGCTTAGG	CCGTCGGATC	AATACAGTGA	TGTCTACTGC	CTTTTTTGAA	3240
GTAACAGATA	TTTTAACGAA	AGAGCAATAT	TTACCGTTAC	TTAAAGAAGA	AGTGAAAAAG	3300
ACTTATGGAA	AAAAATCAAT	GGAAATTGTT	GAAAAAATT	ACCAAGCGAT	TGATGCTACC	3360
TTTGAATCGC	TGCAAGAAAT	TGTAGTTCCA	GAAGATTGGG	CAACCATTGT	TTTAGAGCCA	3420
GAAATCAAG	AGTCGGAaGC	TCCAGCTTTT	GTGACAAATA	TTTTagAACC	AATCAATCGT	3480
CaAGAAGGGG	ATCAGCTGAC	CGTTTCAGAT	TTGATTGAAA	ACGGAATGAA	GGGCGGTGCC	3540
ATTCTATGG	GCACTGCGGC	TTATGAAAAA	CGTGGCATTG	CGTTAGAAGT	GCCTGAGTGG	3600
CAAATGGATA	AGTGTACGAT	GTGTAATGAA	TGTGCCTTTG	TTTGTCCACA	TGCGGCTATT	3660
CGACCGTTTT	TAGCGGATGA	AGAAGAACTG	AAAGAAGCAC	CTGCAGGTTT	TGCCATGCGC	3720
GAAATGCGTG	GCACAGATGG	GTTGATGTAT	CGTATCCAAG	TTCTTTTAGA	AGACTGTACA	3780
GGTTGTGGTC	TTTGTGTTCA	GGCATGTCCC	GTGAAAGACA	AGGCGATTTT	AATGAAACCT	3840
TATGAAACTC	AGAAAGAGCA	AGCGATGAAT	TGGGCCTTTG	CGATGACTCT	TAAACAAAAA	3900
GCCAACCCAG	TCAAAAAACT	TTCTGTAAAG	GGTTCCCAAT	TTAACCAGCC	ATTGATGGAA	3960
TTTTCAGGTG	CTTGTGAAGG	CTGTGGCGAA	ACACCTTATA	TTAAGTTGTT	GACGCAATTA	4020
TTTGGTGATC	GAATGATGAT	TGCTAACGCA	ACAGGCTGTT	CATCAATTTG	GGGTGGTTCG	4080
TCACCGGTGA	CGCCTTATAC	AACGAACGAA	TGTGGCCAAG	GACCAGCTTG	GAGTAATTCTG	4140
CTGTTTGAAG	ATAATGCGGA	ATATGGTTAT	GGTATGTATA	TTGCCAATCG	AACGAAACGT	4200
CAACATTTAG	CTAGTCTGGT	TGAAGAAAGT	CTTGCTAAAA	ACGTCGGATC	AGATTCACTT	4260
CAAGCCTTGT	TAAATGATTG	GTTGGAACAC	ATGGCTGAAG	GGGAAGGAAC	ACAACAGCGA	4320
GCCACCAAAC	TAGCGGCAGC	ATTAAGTGAA	GAAGCCGATG	AAGATCCACT	TTTAACAAAA	4380
ATTTACGAAC	AAAAAGATTT	ACTAGTGAAA	ACGAGTCAAT	GGATTGTCCG	CGGCGATGGT	4440
TGGGCCTATG	ATATCGGGTT	TAGCGGCATT	GATCATGTTC	TTGCCAGTGG	GGAAGATGTT	4500
AATATTTTTG	TTATGGATAA	CGAAGTGTAT	GCCAATACAG	GCGGACAAAC	ATCTAAAGCT	4560

ACACCAGCTT	CAGCCATTGC	CAAATTCTCA	GCGGGTGGGA	AACATACCTC	TAAAAAAGAT	4620
TTAGGGATGA	TGGCAATCAC	TTATGGGAAT	GTCTATGTTG	CACAAGTGGC	TTTAGGCGCA	4680
AACTCAATGC	AAACGATTAA	AGCAATTGAC	GAAGCGGAGC	GGTACCCAGG	CCCATCAATT	4740
ATCATCGGTT	ATACCCCATG	TATTAATCAC	GGGGTCAAAG	GCGGAATGGT	TCGGACGTTA	4800
GATCAAGCCA	AAGAAGCCGT	TGAATCTGGC	TATTGGCCAC	TTTATCGCTA	TAATCCAGAG	4860
TTAGTTAAGA	AAGGCAAAGA	TCCGATGGTC	ATTGACTACA	AGAAAACAGA	TTTGACAAA	4920
ATGCGTGATT	TCCTTGAAAA	ACAAACACGT	TACTCTGCGT	TGCATACAAT	TAAACAGAAT	4980
CAAGAAGTAG	TCGAACATTT	ACTTGATAAA	ACTAAAGAAG	ATGCAATTGA	ACGCTCTGAA	5040
AGTTATAACA	AACTAGTCAC	ACACATGAAA	AAATAAAAAA	ACAAAGCCTC	TTACTTGAGA	5100
GGCTTTGTTT	TTTTATCCAA	TAATAATTTT	TTCTGAAGGG	TATTCGTAGC	CTAAGTCACG	5160
TGTTTCTTTT	GGTACAAAAA	TCATTAAGGT	ATAAAGCATA	CCAATACGTC	CAATAAACAT	5220
TAATAAGGCA	ATCATTAAAT	TACCTACTGT	TGTTAAATCG	TCTGTAATAC	CCAAAGAAAG	5280
ACCGGTTGTC	CCAAAGGCAG	AAGCGACTTC	CACAATGATT	GAAATCAAGG	GCAGATTTTC	5340
AGTGGCTGAC	AAAAAGACGA	CAGCGAAAAA	GCACATAATC	AAAGATAACA	TGAAGACGAC	5400
AATCGATTTT	TTGACATCAT	CGTCATCTAT	TTTACGGCTA	AAGACGCTAA	TATCATCTTC	5460
ACTTTTAAA	AAGGCTAACA	AATACAAGCC	AATAATCGCA	ACAGTCGTTG	TTCGAACACC	5520
ACCACCAACA	GAGCTAGGAC	TACAACCAAT	AAACATTAAC	ACAGAGAAAA	TAATCAACGT	5580
TGTAATTTGG	AAGTCCCCTA	AATCATTAAT	TTGGAGGCCA	GCGTTTCGTG	TAGTCATCGA	5640
ATAAACATC	GAAC TTACCC	AACGGCCACT	TTCATTCAAT	GTCAAAAATA	GGTGATCTTT	5700
TTCTAATAAG	TAAATCAAAA	TAGTCCCACC	AATAAACAGA	ACAATAAACG	CTAGCAAAGC	5760
AATTTTACTA	AACAATGAGA	AGCGGAAAGG	CAAGCCTCGT	TTCTTCTTAC	GAAAGTTCTT	5820
TTTCTTAAAA	TGTAACCACT	CATGAATTTT	CAGTAAAACG	GGAAATCCAA	TCCCGCCGAT	5880
AAAAATCAAG	AACATAATTG	CCAGTAAAAA	GAGATAATCA	TGAGCAAAAG	GAATAATTGA	5940
ATCACCTGAA	ATATCAAAGC	CTGAATTAGT	CACTGCAGAA	ATAGCTTGAT	AAAATCCATA	6000
AAATATCGCA	TCTCGCCAGT	TCGTATAATA	ACCACTTATA	TAAAAATAAA	GCGAAAAGAG	6060
CGAGCCGAAG	ATTACTTGGA	ACCAAATTAA	AATCAAAAAA	GTGGTGCGGA	TCAGTCGAAC	6120
AATCCGCTC	AGTCGCGGCT	GATTCATGTC	AGTCATAATT	AAcTGACGTT	GCTTTAATGT	6180
GATTCGCCGT	TTTGATAAAA	TAAAAAAGAC	CGTTGAAATC	ATCATAATCC	CTAAGCCACC	6240
AACTTGAAAA	AGAACCTCCA	ATAAGACGAT	CCCACGATCA	TTAAAAACCG	AATTAATATC	6300
GACTGTGGTG	AGTCCaGTAA	CACTGACTGT	ACTGATTGCT	AAGAAAAATA	AATCAATAAA	6360
AGGAACATGC	GAGTCAGGTT	CTCTGAAAAT	AGGGAGGCAA	AACAATAAAT	AAGAAATAAC	6420
CGTCATTAAT	ATGTAATAAG	CGACGATGAT	TTGAATAGAT	GAAAAATTTT	TGCGAATAAA	6480
TCGATCGCTt	cTctTTCGCT	GTGCCAGCCA	CAGCAAACGT	GTATGGTTCA	CTAGATaAcy	6540

TCCC <sub>a</sub> ATTCT AATTAACTA mGTATACCA AACTTTTCGC CTCAAACGAA TAATTCTGCC	6600
AGAAAATCCT TGAAATGACA AATAAAAATA AGAAATTCGT TAGAAATAAT TGGAAAACAT	6660
CCACAAACAA GCGAGAACAT GATATTTTAT GAAAGAATAG AAAAATAGGT TGTCCCCAA	6720
ACAACTTATT AATAATTGGT CAGTTTGTCT AACGATACAA CTTTAAAAAG TAGTGGTAGC	6780
TTTTTTATTT CGGTTGATAT GTGAATGTGT TCACTTTATG TAACAACTG TTATTTTAT	6840
TCAATGAAA AATAAGAAAT TGTGTTGGTA GAAATCAATA GTTTATAAGG TGGGAAAACA	6900
ATGAAAAAAG TGTTAATGGG AGTTCTTTCC CTAGGGTTAT TATTAGGTGC AGCAACAGGT	6960
TGTACAAGCG ATCAAGAAAA AGCCGCTGGT AAAACAAAAG CGTCTTCTGA AAAACAGAA	7020
GCAACATCTG GTGCCTCTGC TAATGGTTAT ACAGATCCAA GTGAACTAAA AGACAGTTAT	7080
GATGTTGTTA TTGTTGGTTC AGGTGGTGCG GGGATGACCG CTGCATTACA AGCCAAAGAA	7140
GCTGGCATGA ATCCAGTTAT TTTAGAAAAA ATGCCTGTTG CAGGTGGGAA TACAATTAAA	7200
TCGTCTTCAG GAATGAATGC TTCACAAACA AAATTCCAAG AAAAGAAGG CATTAAAGAT	7260
AGCAACGATA AATTCTTTGA AGAAACACTA AAAGGCGGTA AAGGAACCAA CGATCAAGAA	7320
TTACTACGTT ATTTTGTTGA TCATTAGCT GAAGCAATTG ATTGGTTAGA TACAAAAGGr	7380
ATTACATTGA GCAATTTAAC GATTACAGGT GGAATGAGCG AAAACGCAC ACACCGACCT	7440
GCTGACGGAT CAGCAATTGG GGGCTATCTA GTTGACGGTT TAGTACGTAA CGTTCGCGAA	7500
GAAAAAATTC CATTGTTTGT TGATGCTGAC GTCACAGACT TAGTTGAAG <sub>a</sub> AAACGGTC <sub>a</sub> A	7560
ATCGATGGCG TAAAAGTGAA AATGAAAGAT GATAAAGAAA AAACAGTTAA AGCAAAAGCA	7620
GTCGTTGTCA CAACAGGTGG TTTCGGTGCC AcGAAAAATT AATTACACAA TACAAACCAG	7680
AATTGAAAA CTATGTTACA ACGAACCAAG AAGGCACAAC AGGTGACGGT ATCCAAATGA	7740
TTCAAAAAGT TGGcGGTGCG TTagTGGATA TGAAAGAAAT TCAAATTCAC CCAACTGTTC	7800
AACAAAGCGA TGCTTTCTTA ATTGGTGAAG CAGTTCGTGG AGAAGGCGCA ATCTTGGCTT	7860
CTCAAAAAGG AGAACGTTTT GTTAATGAGT TAGATACACG CGACAAAGTT TCTGCAGCAA	7920
TCAATGCCTT ACCAGAAAA TCAGCTTATT TAGTCTTTGA CCAAGGCGTT CGCGACCGTG	7980
CAAAAGCCAT TGATTTCTAT GATCAAAAAG GTTTCGTGGA AAAAGGTGAA ACGATTGAAG	8040
AATTAGCTGA AAAAATTGGC ATGCTGCGG ATACATTGAA AGCAACGATT GATACTTGGA	8100
ACCAAGATGT TAATGCCAAA GATGACAAAC AATTTGGCCG GACAACTGGA ATGGAAGCTG	8160
ACTTAAGCAC AGCCCCTTAC TACGCAATTA AAATTGCACC AGGGATTAC CACACAATGG	8220
GTGGCGTGAA AATCAACACG AAAACAGAAG TTTTACGTGA AGATGGCACA CCAATTAAAG	8280
GTCTATATGC TGCTGGTGAA TTAACAGGTG GCTTACATGG TC <sub>a</sub> AAACCGT ATCGGAGGCA	8340
ATGCAATCGC CGATATTATC ATTTATGGTC GTCAAG <sub>c</sub> AGG TACGCAATCA GCAGAATTTG	8400
CCTCTGCTCA AAAAT <sub>n</sub> ACTA GCATTAAAA GTCTGAAATT GTAGTCAAAT AACTACAATT	8460
TCGGGCTTTT TTATTGCATA CTAGCAGGAT CATTAGTTGG AGTATAAATA GCCCTACTAA	8520

AATGACTTTG	GCGTAGATTC	TGTTTTTTTC	AGAAAAATGC	ACGACCATTT	TAGCTAAAAA	8580
TACTGCTTTT	TTGCTGAAAA	TATAGTAAAA	TACGTAAGGA	ACTAACGGAA	AGGTGTAGTG	8640
AACATGGCTC	AATTATTTTT	CAAGTATGGC	GCAATGAACA	GTGGTAAAAC	CATTGAAATC	8700
TTAAAAGTAG	CACACAATTA	CGAAGAACAA	GACAAACCAG	TTGTTTTGAT	GACCAGTGGA	8760
CTAGATACAC	GTGATGGTGT	TGGGAAAGTA	TCCAGTCGCA	TCGGATTACG	CAGAGATGCT	8820
ATTCCAATTT	TTGAAGAAAC	CAATGTCTTT	GACTTGATTA	ATGATTTGTC	TTACAAACCT	8880
TTTTGCGTAT	TAGTCGATGA	ATGTCAATTT	TTAAACAAGC	ATCATGTCAT	TGAATTTGCA	8940
CGAATTGTTG	ACGAATTAGA	TATTCCAGTA	AtGGCATTTG	GGcTAAAGAA	TGATTTTAGA	9000
AATGAATTAT	TTGAAGGCTC	AAAGTATTTA	CTTTTATATG	CAGACAAATT	AGaAGaATTA	9060
AAAACGATTT	GCTGGTTCTG	TCATAAAAAA	GCAACTATGA	ACTTGCACTA	TATTGATGGC	9120
AACCCCGTTT	ATGAAGGGGA	TCAAGTCCAA	ATTGGTGCGA	ATGAAGCCTA	CTATCCTGTT	9180
TGTCGCAAAC	ATTATTTTCA	TCCAAAAACA	GTGAATGAAG	AACAGTAAGG	AGAGACACTA	9240
TGTACGATCA	ATTwCAATCT	ATTGAAGATC	GCTATGAGGA	ATTGGGCGAA	TTATTAAGTG	9300
ACCCAGCAGT	CATTAGTGAT	ACAAAACGTT	TCATGGAATT	ATCGAAGGAA	GAAGCTAATA	9360
CACGTGAGAC	GGTTGAGGTT	TATCGCCGAT	ATAAACAAGT	TGTTGAAGGG	ATTAGCGATG	9420
CCGAAGAACT	TTTAAGTGAA	AACTTAGATG	CGGAAATGGC	CGAAATGGCG	AAAGAAGAAC	9480
TTTCTGATTT	GAAAAAAGAA	AAAGAAGTGT	TAGAAGACCG	CATTAAAATT	TTATTATTAC	9540
CAAAAGATCC	AAATGATGAT	AAAAACATCA	TCATGGAAAT	CCGTGGTGCA	GCCGGTGGCG	9600
ATGAAGCCGC	ACTTTTTGCT	GGCGACTTAT	TTAATATGTA	CCAAAAATAT	GCAGAAGCAC	9660
AAGGGTGGA	AGCGGAAGTT	TTAGAAGCGA	ACGTCACTGG	AATCGGTGGA	TATAAAGAAG	9720
TAATTATGAT	GATTTTCAGGA	GATAACGTCT	TTTCAAATT	AAAATATGAA	AGTGGCGCTC	9780
ACCGCGTGCA	ACGAGTTCCT	TCAACAGAAT	CACmAGGACG	GATTCATACA	TCGACAGCAA	9840
CCGTTGTAGT	TATGCCGGAA	GCGGAAGAAG	TCGAAATCGA	ATTAGCTGAT	AAAGATATCC	9900
GTGTCGATAT	TTATCATGCA	AGTGGCGCTG	GTGGACAGCA	CGTCAATAAA	ACCGCTTCGG	9960
CTGTTTCGATT	AACCCATTTA	CCAACAGGAA	TCGTAGTCGC	AATGCAAGAT	GAACGTTTAC	10020
AGCTAAAAAA	CCGTGAAAAA	GCAATGAAAG	TTTTGCGTGC	TCGTGTGTAT	GACCAAATTC	10080
AACAAGAAGC	ACAAAGTGAA	TATGATGCTA	ACCGTAAATC	GGCGGTGGGG	ACCGGAGATC	10140
GTTCAGAACG	AATCAGAACG	TATAACTTCC	CGCAAATCG	AGTAACGGAT	CATCGTATCG	10200
GTTTAACCAT	CCAAAACTA	GACCAGATTC	TTGCTGGCAA	ATTAGATGAA	ATCATTGATG	10260
CATTAGTCTT	ATATGATCAA	ACATCGAAAT	TAGAAGAGAT	GCAAAATGGC	TAAACGTTAC	10320
TTTGAAGTCC	TTAACTGGGC	TTCTTCTTTT	TTAGAGGCGC	AAGGCAAAGA	AGGCTACGCT	10380
ATTCACTATG	TTTTTTTAGA	ACGTAAAGGT	TGGGATAAAA	CGCAGTGgCT	CCTGCATATG	10440
CAAGAAGAAA	TGCCCCAAGA	AGAAGAGGAA	CAATTAAAAA	CGGATTTAGC	CCAATTATTG	10500

ACAGATTACC	CAGCACAATA	TTTACTAGGA	CAAGCTGAAT	TTTATGGACA	TTCCTTTATT	10560
GTTAATGAAC	ATACTTTGAT	TCCTAGGCCA	GAAACAGAAG	AATTAGTCGA	GCGTTGCTTA	10620
AAAGCCAATC	CAGATACTCC	TTTGACGGTC	GTGGATGTCG	GAACAGGGAC	AGGTGCCATT	10680
GCGGTTAGCT	TGAAACTGGC	ACGGCCCAAC	TGGCGAGTTA	TTGCCATTGA	TCTTTCAGAG	10740
GAAGCGTTGA	CTGTCGCTAA	ACAAAACGCG	CAAGCATTAG	GTGCTGGCAT	CGAATTTTAT	10800
CATGGGAATG	GCTTACAGCC	AGTTGCTTCT	GAAAAAATTG	ATCTGCTTAT	TTCTAATCCG	10860
CCCTATATTA	GTGAACAAGA	ATGGTATTTA	ATGGATGCAA	GTGTCCGAAC	GTATGAACCT	10920
AAAACCGCTT	TATTCGCTGA	AAATAACGGT	TTGGCTTTAT	ACCAACAATT	AATTCATGAA	10980
AGTCAAACGA	TGCTTAAAGC	AGACGGTAAA	ATTTATTTTG	AAATAGGGTT	TCAACAAGGC	11040
GCGGCCCTCC	AAGAATTATT	GAGCGCCGCC	TATCCACAAA	AAACAATTAA	AATTGAAAAA	11100
GATTTATCTG	GCAACGATCG	CCTCGCGATA	GCCGAGTAAC	TAAGGTTCTT	CGTCGCCACT	11160
TTTCTAGAAG	GAGCGATTTG	CAATGGAAAC	AAAAAAATTA	ACCAAACAAG	AAATAGCCCA	11220
AGCAGCAGCA	GCCATTCAAG	CAGGCGAGCT	AGTGGCTTTT	CCAACAGAAA	CTGTTTATGG	11280
CTTAGGTGCC	GATGCCTTAA	ATGAAGCGGC	AGTAAAAAAA	GTTTACCAAG	TAAAAGCCG	11340
CCCCAGTGAT	AATCCATTGA	TTGTTTATGT	GAACAACGTT	GAGATGGTCA	AAAAATACGT	11400
GGCGACCTTA	CCAGAAGTCG	CTACAAAATT	GGTGTCTGCC	TTTTGGCCGG	GACCATTAAC	11460
TTTAATTTTT	AATGTGCCTA	GCGATACATT	TTCTAAAACA	GTAACAGGTG	GTTTAGAGAC	11520
AGTCGCTTTT	CGTATGCCAG	ATAACCAAGC	GACCTTAGCT	TTGATTGAAC	AAGCCAATGT	11580
TCCGTTGGTC	GGACCAAGTG	CGAACACTTC	TGGCAAGCCT	AGTCCAACAT	CTGCAGAGCA	11640
TGTTTATCAT	GATTTGCAGG	GGAAAATTGC	TGCCATATTA	GATGATGGCC	CAACGCAGAT	11700
TGGGGTGGAA	TCAACGGTCT	TGGATTTAAC	GGCGGAAGAT	GGTGTGCCAG	TTATTTTACG	11760
ACCAGGAGCA	ATTACCAAAG	AGCAGTTGGA	AAGCATTATT	CCCCAAGTGA	AAATCGATAC	11820
ACATTTGATT	AGTGAACTG	CAGCACCCAA	AGCGCCAGGG	ATGAAATACA	AACACTATTC	11880
TCCTGATGCG	GAAGTTTGGA	TTATAAGCGG	TCAAACAGCA	GAATGGCAAG	CGGCGATTCA	11940
GCAAGCAAAA	GAACAACAGG	AAAAAATCGG	CTTATTTTTA	AGTGATGAAC	AGGCAGCACA	12000
ACTAGATACA	GAAAATGCGT	TTGTTTATTC	TTATGGTGCA	GAAACGGTTG	AAAATGCCAC	12060
AAAAGAACTA	TTTGCAGGGT	TAAGAGCGTT	AGATGAGCAA	GGAGCTACCA	CGATTTTTCG	12120
GCAA <sub>g</sub> GATTT	GCAGAACTG	GTTTAGGCAC	TGCTTACATG	AATCGCTTGA	AAAA <sub>a</sub> TCAGC	12180
CAATCAAAAA	TTTTTTGAAA	AATAAGAGGT	ATTTGTAACG	AAGCTGACAC	AGAAGTATGA	12240
TACAATCAGA	CAAAAGAAAA	CAAAATGGAG	TGTTGTTTGA	ATGG <sub>m</sub> TTACA	AAACGTATGA	12300
CCCAGATTTA	TGGAATGCAA	TTGCAAGAGA	AGAAGAGCGC	CAAGAAAATA	ACTTGGA <sub>ACT</sub>	12360
AATCGCATCT	GAGAATGTCG	TGTCAAAAGC	AGTTATGGCT	GCCCAAGGAA	GTATTTTAAC	12420
GAATAAATAC	GCAGAAGGTT	ACCCTGGCAA	ACGGTACTAT	GGTGGTTGTG	AATTTATCGA	12480



TATCGTAGAA AATTTAGCTA TCGATCGTGC CAAAGAATTA TTTGGTGCAA AATTCGCGAA	12540
TGTACAAGCC CATTACAGGT CTCAAGCCAA TACAGCGGCA TACCTTTCAT TGGTTGAACC	12600
AGGTGATACC ATTTTGGGGA TGGATTTATC AGCTGGTGGT CACTTAACAC ATGGTTCGCC	12660
CGTTAACTTT AGTGGAAAAA CCTATAATTT TGTCAGTTAT GGAGTGGATC CTTCAACAGA	12720
AGTAATCGAT TACGATGTCG TGCGAATTTT AGCAAGAGAA CATCGTCCAA AACTAATCGT	12780
TGCAGGCGCA AGTGCGTATT CACGAACGAT TGACTTCAAA CGTTTCCGTG AAATCGCTGA	12840
TGAAGTAGAT GCCAAGTTAA TGGTTGATAT GGCGCATATT GCAGGCTTGG TAGCTTCaGG	12900
GTTGCaCCCA AATCCaGTTT CgTATGCTGA TATCgTAACA AGTACgACCC ATAAAAcATT	12960
GCGTGGTCCC TCGTGGCGGT TTAATTTTGA CAAATAGCGA AGAATTGGCG AAAAAAGTAA	13020
ATAGTAGTAT TTTCCAGGC ATTCAAGGTG GCCCATTGGA ACATGTGATT GCCGAAAAG	13080
CCGCAGCTTT TAAAGAAGCA TTGGATCCAA GTTTTGCTGA ATATAGTCAG CAAGTGATTG	13140
CAAATGCGCA AGCGATGACG AAAGTTTTCA ATCAAGCGCC AGAAGCTCGT TTAATTAGTG	13200
GGGCAACAGA TAACCACTTA TTAGTGATTG AAGTGACAGG TTTTGGTTTA AATGGAAAAG	13260
AAGCAGAAGC TATTTTAGAT AGTGTAATA TTAGTGCAA TAAAAATTCA ATTCCGTTTG	13320
AACAATTAAG TCCATTTAAA ACAAGTGGTA TCCGTATTGG AACGCCTGCA ATTACTTCAC	13380
GTGGTTTTAA AGAAGAAGAT GCGGTAGAAG TTGCTAAATT AATCGTTCAA GTCTTAAAAG	13440
ATCCAGAAAA TACAGCAGTT CATGATGAAG TCAAAGCAGC TGTAGCAGCG CTAATAAAA	13500
AATATCCATT ATATACTAA CATCAGGGGA TAGCTGAAGC GCCGATTCTC TGGTACAAAT	13560
AAGAGTCGGA GCTT	13574

## (2) INFORMATION FOR SEQ ID NO: 115:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4528 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## .(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

ATTTACAGCG TTAGTGAhAC AAGCAAAAGA CGCTTTAAGC AAATAAGCTT AAACAATCTT	60
TCACACCCCTA CTCAAGGAGT AGGGTGTTTT TTTAATAACT CCGTTGGTAA CTTTCATCTA	120
TAAATGAAAA AACTTGTGAG AGCGGAATTT CTTAGAAAG ACTTAACCAA TGCTCTTTAT	180
TCATATGATA GGCAGGAAAA TAGCCAGGTT GGTCTTTTAA TAAGGCAATC AATTCAGGCG	240
CAAGTTTAAC GGTTAAAATA GTGAGCTGCT TATTCTCGGT AGCGCCGAGC AGTTGTGCAT	300
CCAGTGGCAT AATTAAAGCA AACCATTTTG CATTTCTAGG ATGACGAAAG ACGAGATAGT	360
CTGGGTGTTT TTTAAAAGGA AAATCGGGCG TTATCTGATA ACGTCTTGG ATATAGGCGC	420
TAATAGTTTC TTGAAAAGTC ATAGGTTTCT TCCTTTCGTT AAGGCTAGGC ATATCTATTT	480
ACATCATAGC AAAAAATTACA TAAAATTACT TCGTTCAAAG GAATGGGAGG AGGGAGCAGA	540

TGACATTAAT	TTCAACAATT	TTAGTAACCT	TGGTGGCTGT	AGAATTTTTC	TACATTATGT	600
ATTTAGAAAC	GATTGTACCA	ACATCTGAGA	CAACAAGCCG	TGTTTTTAAA	ATGGATAGGC	660
AAGAGTTACA	AAGArAATCA	GTCACCTACAT	TGTTTTAAAAa	TCAaGGAATC	TATAATGGAC	720
TAATCGGTGC	AGGACTTATT	TATAGTGTCT	ATTTTGC GCA	AgCAACGATG	GAGATGACAA	780
AATTTTTTATT	GATTTATATT	ATTTTAGTGG	CGGCGTACGG	TAGCTTAACT	AGTGATAAAA	840
AAATCATTTT	AACACAAGGC	GGTTTAGCTA	TCGTGGCTTT	GATTAGTTTA	CTTGTTTAGC	900
CATCAGAGTT	AATAGAGAGA	GGCAGTCAAA	TCGACAGAAA	AAACACGGTT	GATTTGACTA	960
TTTTTTATAG	AATTGAGAGA	GACAAAATGT	GGTATACTTT	GTTTATTAGA	TAGGGAGGAA	1020
TTAAAAATGC	TGAACCAAAA	AGAAGTAAAA	AAAACCGATT	ATACTCAATT	AATTTCTCTA	1080
TGGCGTTCTT	CAGTAGAAGC	TACCCATTTG	TTTTTGAGTC	AGGCAGATAT	TGATAAAAATC	1140
GAAGTGGTTT	TACCTGATTA	TTTTCAACAA	GTTCAATTAA	GTATGTGGTT	AAATGAAGAA	1200
CAGAAaTGTG	TGGGCTTTAG	TGGAACAAAT	CAACAAaCGT	TGGAAATGTT	ATTTATTGAC	1260
CCAGTCTATT	TTCGTAAAGG	CTATGGTGGa	GAAATTATCC	mAAAGTTAAT	AGAGCAAGAA	1320
TCAATTATCT	TTGTTGATGC	TAATAAACAA	AATGAAGGGG	CAGTGAAGTT	CTATCAATCT	1380
CAAGGGTTTC	AAGTAATTGG	TGAATCAAAG	GAAGATCCTC	AAGGAAATCC	GTTTCCTATT	1440
TTACATATGA	AACGGATATA	GAAAAGAAGG	GAGAAGCCAT	GCGAGTAATC	GTTACAGAAT	1500
ACCAACCAGC	CTGGGTGGAA	CAGTTTGAAG	AAGAAGCCCA	AGCGTTGAAA	CAGATTCTAA	1560
AGGAAAATTG	TCTTAAAGTT	GAACATATCG	GTAGCACATC	TGTGcTAAt	TTAGCAGCCA	1620
AACCAATTAT	TGATTTTTTG	GTCATTGTTG	AAGAAATTGA	AAAAGTAGAC	CTGTTACAAT	1680
GGGAATTTGA	AAGAATCGGT	TATGAATATA	TGGGGGAATT	TGGGCTATCA	GGACGTCGTT	1740
ATTTACGAAA	AGGCCCGATA	AAAAGAACAC	ATCATGTGCA	TATTTATCAG	TTTGACAATA	1800
CGCAAGAAAT	TTTGCGCCAT	CTTGCTTTTC	GAAATTATTT	ACGAGAAAAT	CCGGCAATTG	1860
CTACAACCTA	TGGTACATTA	AAAAAGCAAT	TGGCCCAAGC	ACACCCCGAT	AGCATTGATA	1920
AATATATGGA	TGGCAAAGAT	GCGTTCATTA	AGAAAATAGA	AAAAGAAGCG	TTGAAGAAAT	1980
ATTGGGAAAA	ATAAAAAAAG	TTGCTAAGGC	TAGCAACTTT	TTTTTGGTTC	TTTTTTGAAA	2040
AAATGATAAA	GAAGACCACT	TAAAAGAACA	CCTAGAATAA	TTCCAATCGC	TAAATTATGC	2100
GTATAGACGA	TCACACCCAC	TGTCACGAGT	AAAATAACAA	TTTCAGTAAT	TTCAAAAAGTT	2160
CGGAAAAGCT	GCAAACCTTC	CCAATCAAAG	GTGTCGACTG	CCACTGTCAT	CATGATCCCA	2220
ATCAATGCAG	CTGTAGGAAT	GGCTACCATC	ACGGATTTAA	AAACAAAAAT	GAATAGAAGG	2280
AGGGTAATAC	CTGAAATCAA	AGTAGAAAGC	CGTTTTCGTC	CTCCAGATTT	AACATTAATG	2340
ACCGCTTGCG	CAATCATCGC	ACAGCCAGCT	TGGCCACCAA	AAAAGCCGGT	AATTATATTA	2400
GCGAATCCTT	GTGCTTTAAC	TTCTCGTTGG	CTATCACTTT	GGCTCGCTGT	CATTTTGTCC	2460
ACAATGGGGA	TCGTCAATAA	AGATTCAATT	AACCCGACCA	TTGCTAAGGC	AACAGAGCTA	2520

GGTAAATATA	TCCACAACGT	TTCTAAGTTG	ACGGGTACAT	TAGGCCAACG	GATCGCTGGA	2580
TGAAACTGAC	TCATTTACC	AAGATCACCT	ACCGTTTGCA	AATGACCTTT	TAAAAAATAG	2640
GACAGGATAG	TCATGAATAA	AATAACAATT	AAAGCAGGAG	GCACAATTGT	AATAAACTTA	2700
GGTAATAGAT	AAATAAGTAA	GATACTACAG	ATAATCATAC	CGTAGGACCA	CAGTGTTTGA	2760
TGGGGCAATT	GTGTACTTG	CGCCATGAAA	ATTAAATGG	tAACGCATTG	ACGAATCCGT	2820
ACATCACTGT	CTTGGAATA	AAGCGCATCA	ATTTGTGGAT	ACCTAAATAA	CCTAAAATAA	2880
ATTGTAGAAA	GCCTGTTAAA	ATAGTCGCAG	CAAGCATGTA	TGACAAACCA	TGTGCTTGAA	2940
TTAAGCCTGC	aAGGACTAAG	GCCATTGAGC	CAGCAGACGC	CGAAACCATT	GCCGgACGGC	3000
CGCCCGTAAA	AGTAATCACA	AGTAAAGTAA	cAGctGAAGC	AAACACTGCA	GAAAGCGGGT	3060
GAACACCAGC	AATAATCGCA	AACCCGATTA	CTTCTGGTAA	AATGGCGACA	GAAGAAACGA	3120
GACCGGCAAA	GAGATCATTT	CGTTCATTGC	CTAACCATTTC	GCTTTTTTTA	ATACTAAATA	3180
ACATCGAAAA	ACTCCTTTAA	GTTAATAAAA	CCTTATTCAT	AATAGCATGA	ATAAGCTGAA	3240
ATGAGCGTGG	GACAAAAATC	ACTTTGGATT	TTTGCTCCaC	GCTCaAAAAC	TGATAAACGG	3300
CGGGAACAAA	AGCAACTCCT	TCGGAAATAA	GCCGGAATTC	TCCAAAAATT	ACAGAACAAT	3360
TTTCGGAAAT	TCCTTCTTAT	TTCTCGGAGT	TAAACGCTTC	TGTCCCGACC	TCATTAGAAC	3420
TGTAGCAGTG	CCAACAACAA	AATGGCTAGG	ACGACTTGAA	CAGTTCCTAC	TGATAAACCA	3480
TAGATTAAGA	AACGTTT3CC	TTCTTGGAAG	AACTTTTTTAA	AATCGAGTCG	TAACCCGATT	3540
GCTGCCAAGG	CAGTAATTTT	AAACCAAGAA	CTAAAGAAAT	GAGCAGTCTC	ACTGATCACG	3600
ACGGGGAAAT	GAATCAAAC	ATTAAAGACA	CAGGCAATGA	AAAAGCCAAC	TACATACCAA	3660
GGTAGGGCGC	TGCTTTTTTT	GGTGACTTCT	ACCAACTCAG	CCTCTGATTC	TGCCGTCTTA	3720
CTTTGCTTGA	AACGTCCAAA	TAAATAGACA	ACTGCTACGA	GTAGGACAAT	CCGCATAATT	3780
TTAAAGAGCA	TCGCAAGTTG	AaTGCGTTC	TCATTGACCA	TATTGGCACT	GGCGACAACC	3840
TGACCAACTG	ATTGTAAAGT	CCCTCCGATT	AGTGCGCTAC	GAGCCAACAA	ATTGGTACCA	3900
TAGAGAATCC	CACTTAAAT	GGGTAAGGTC	AACATTAAAA	CGGTTCTTAA	TAAGTTTACT	3960
AAAGTAATGA	TTTGGCCTTT	TTCTTCTTCA	TCTGCTTGAA	TTGCTGGAGC	AATCGAAGCA	4020
ATAGCAGAAG	ACCCACAAAC	GGCATTACCA	CCAGCCATCA	ATAAGACAT	ATTATCAGAA	4080
AAAGCCAGTT	TTTTACCGAT	CAAATAGGCA	AAAATAATGG	TCAGCGACAT	TTGAATGAGA	4140
ATAAATGCGA	CACCTTGTA	GCCAATTTGG	GCAATCGTTT	GAAAGGTCAC	GGTGGTTCCC	4200
AATAAAACAA	CAGAAAATTC	TAGTAATTTA	CTTTCGGCTA	CTTTGGTTCC	GCGATTTAAA	4260
TTGGCACCGC	GAACAAAGGT	ATTTCCCAAA	AAGATTCCTA	GTAATAATAGC	TATCGTAGCG	4320
GCGCCTAAGC	TAGGTAACCA	GATGGCAAGA	AACTTGCTAA	GGCAGGCGAC	AAGGAAAGCT	4380
GTTAGCAAGC	CGGGAAGAAT	TTGGAATAAA	GATTGGATAT	AATCATTTTC	TGAATTTTTC	4440
ATTGAAAAAA	TCTCTCTTTC	TTGTTTATTA	CTTCCAGTAT	AACGAAATTT	TTTACATTTG	4500

AGAnATAAAAT AGTTAAATA ATATTTAT

4528

## (2) INFORMATION FOR SEQ ID NO: 116:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5147 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

GAAGTCAGTG AAGCGAAATT AAAAGAAGCA GCCAATCTTA TGCCAGAAGA GCCGTTAATT	60
CAATTTGCTT TGGGAGAACT ATACTTTACA AATGGTCAAT TTGTCGAGGC TATTACTCGT	120
TATCAAAGCA TTGTAGAAAG TGGGACAGCG CAAATCTCTG CTATTTTCCTT GAATGAACGC	180
TTAGGAAGTA GTTACAGCAT GTTAGGTGAT TTTGAAGAAG CGGTTCCCTTA TTTGGAAGCC	240
GCAGTCAAAG AAGAACAAAC AGACGATCGT TTATTCCAGT TAGCTTTCAC TTATTTACAG	300
TTACATGAGA ATCAAAAAGC CATTGCGTTG TTCCAGCAAT TAAAAAACT AAACCCACAT	360
TATCAATCAC TTTATCTATA TTTAGCGGAA GCGCTCCAAG AAGAAGAGCA ATTAGAGGAA	420
GCAAAAGCGG TGATCGAAGA AGGAATTGCG GAAAATCCCT TCCAAGTAGA GTTATACCAA	480
TTTGCCTCTG AAAATGCCTA TCGCTTACAC GATATTGCTG GTGCTGAAAA CTGGTTGTTA	540
AAAGCTTTGG AATTAGGTGA GAAAACAGAT GAAAGTCGTT TGACCTTAAG TAATCTCTAC	600
ATCACGGAAG AGCGTTTTGA GGAAGTAATT ACTATCTTAA ATGATTTAGA GGAAACGGAT	660
CATCCTTATG CTGAATGGAA CTTAGCAAAA GCGTACAACG AATTAGAAGA CTTTGCCGTC	720
GCTAAAGTTC ATTATGAACA AGCAGCGCAA GAATTGTCTC ACGAACCAGA ATTTCTAAAA	780
GAATACGCAT TGTTCTTACG AGAAGAAGGC GAACTTGAAA AAGCAAAACA CTTGTTATAC	840
CACTATTTGG AACATGAACC TGGTGATGTC GACATGCAAT CTGTGCTAGC AGATATGGAA	900
GAAAGATAGG AAACGACGAT GTTTGTTAGT TTATCAGAAA AGAAAAAATT TTTAACTTGG	960
TTAGTGAATA CCGCGCCTTT TGGAAGAAGA GAAGTACTGT GGATTTTAAA TTACTTATTA	1020
ACGCACGATG CGATTTTAAA TAATGTTTAT TTTGTTGAGA ACGTAGAAAA AACCGATCGG	1080
GGGATTCGTG TGGTAGCGGA TGGTTTAGGT AAGGAGCCGT TATTGCTTTT TATCCAAGCA	1140
CAAGAATTTA CAGATCCGGA ACAAATTTTT CACGAGATTC GGATGAATTG GCGGAAAGCT	1200
CTTTATTTAG AATGTGTATT TCCAGAGGCT TGGCAAACGA GCCAGTATTT ATCTGTTTTA	1260
GAAGATAATC CGTTTGCGCC ATGGAATGAA CAAGTGGATC AGGAAGTTGC CAGAGCGATT	1320
GATCAATATT TTAAGCAAGA AGAACAAACG CAGCGAATGG CATTACTTAA AGCTCAAATT	1380
GATGATGCTT TAGAAACAGG CAACAAAGAA GCGTTTTTAG AATTATCCGA TGAATTAAAT	1440
CGTTTGAAAC AACAAATAAT AAAACCCAGC ATTTATTGAA AATGCTGGGT TTTATTTATT	1500
GTTTTAATTT AAAGCGCCGT TGTTTGCTGG GTTTTGCATA AGTAAAGCCT TCGCCAATTG	1560

CTTCATGTAC	GTTAATGATG	GACACAAAGG	CTTTTTCATC	TATTTTCATTA	GCAATCCGTT	1620
TGACTTCCAT	TATTTCTCGA	GGGCTGACTA	CCGTGTAGAG	AACCCGTTTT	TCCACTTGTG	1680
AATAAGCACC	TTCACCATTT	AAATATGTTA	CCCCACGTTT	TAAAGAAGCC	ATTAACTTT	1740
GACCGATTTT	TTCGGAATGA	TTGGAAACAA	TTAAAATTCC	TTTCGCAGCA	TAGGCTCCAT	1800
CGAGTACAGA	ATCAACAACA	CGACTGAAAA	CAAAGGACAC	AATTAAAGTA	TACATCATTC	1860
GTTTGACATC	AATATAGGTC	AACGAGAGGA	GTAAGACTGC	AACATCAAAA	AGTAACAACG	1920
AGCGGCCCAT	GCTAATCCCA	AAGTATTTTT	CGAAAATGCG	GGCAATGACG	TCGGTCCCAC	1980
CAGTGGTTCC	GCCTACGCGG	TAAACAAGGC	CACTACCGAA	TCCTGCGGCC	AAACCAGCTA	2040
AGAGAGCAGC	AATCAATAAA	TCATGATCTA	AATTAATTTT	AACAGGAAAT	TTTTGCCAAA	2100
ACCACAAAAA	TACCGATAAC	GACACGGTAC	CAATTAGCGT	ATAATAGAGG	GAACGTTTGC	2160
CTAAAATTTT	CCCACCAATT	AATAAGAGCG	GAATATTGAT	TAATAATGTC	GTATATGCTG	2220
GGTTAAAATG	AAATAAGGCT	CTTAAAATCA	AAGTAATCCC	CGTGACACCA	CcTTCGGCCA	2280
GGTCATTGGC	AATATTGAAA	GTAACCAAAC	CAAAGGCATA	GATACAGCTT	CCTAATACAA	2340
TAAATAACAC	ATCTTTAAAA	TAAAATCTTT	GTTCTTCCAC	TTAATCACGT	CCTTTTACAG	2400
TATAAACTGT	ATCATTCTCT	ATATAAAACA	ACAAGTAGAA	CCTTTCCAGA	TAGCCAAGTT	2460
TTTGATAAGA	TGTAGGAGGT	AGAGTAAGAA	AGGTTGGGGT	AAAATGTCAG	AATCAGAACA	2520
ACGTAGTTTG	CAATCGATGC	AAACAGAAAT	TGATGATTAT	ATTCAACAGT	TTAAAACAGG	2580
CTATTTTTCA	CCATTAGGAC	AAATGGCTCG	TTTGACGGAA	GAGGTCGGAG	AATTAGCCCG	2640
TGAAGTCAAT	CATTATTATG	GTGAAAAACA	GAAAAAGCA	GACGAAAAAC	CAAAGACAGT	2700
ATCAGAGGAA	CTTGAGATG	TCTTTTTTGT	ATTGATTAGC	ATGGCAAATT	CTTTAGAGAT	2760
TGATTTAACC	GAAGTTTTTG	AAGAAATAT	GGCCAAATTT	AACAAGCGGG	ATCGCTATCG	2820
CTTTGAAAGA	AAGGATGGGA	AAACACATGA	TTAAAATAAT	TGTCGCTGGG	TTTAAAGGAC	2880
GGATGGGCAG	TACTGCTACA	CAAATGGTCT	TAGAAACCGC	AGACTTTGAA	TTAGTAGGTG	2940
TGTATGATCC	TCATGAAGCA	CAAGAACTG	TTTCTTTCAA	TGATGAAACA	GCAATCCCCG	3000
TTTTTCAGCG	CTTAGAAGAA	GTCCTAGCTG	TCAAACCAGA	TGTTTGGATT	GATTTTACGG	3060
TTCTGAAGC	CGCCTATCCA	AATACGCGTT	TTGCGTTGGA	ACACGGTATG	GCGCCTGTTG	3120
TTGGGACGAC	AGGCTTTACG	GAGGAACAAA	TCAACGAATT	GACAAACCTA	TCTCGAGAAA	3180
AAGCCATCGG	CGGCTTGATT	GCGCCAAATT	TTGCAATTGG	GGcAGTTTTA	ATGATGCAAT	3240
TTGCCCAAAA	AGCGGCGCAG	TATTTTCCAG	ATGTGGAAAT	CATTGAATTA	CATCATGATA	3300
ACAAATTGGA	TGCACCGAGC	GGAACGGCTA	TAAAACTGC	AGAAATGATT	CAAGAAGTTC	3360
GGCCAGCAAA	AAAACAAGGG	AAcCGCAAGA	GGTAGAATCA	ATACCAGGGG	CACGCGGGGC	3420
TGATTTTCAG	GGTCTGCGAA	TCCATAGTGT	CCGCTTACCT	GGTTTAGTGG	CACACCAACA	3480
AGTCCAATTT	GGCAGTGTAG	GTGAAGGGCT	CACTATTCGG	CATGATTCTT	ATGATCGGCG	3540

TTCATTTCATG ACAGGAGTAG CGTTAGCTTG tCGGCAAGTG GTACAGAGAA CAGAACTACT	3600
TTACGGATTG GAACAGATGC TATGAAATTG ACAACTATTC CAAACGAATT TAAAGAAGCC	3660
GCTCCCGTTA TTCGTGAAAT CAATGCACAG GGCTTTGAAG CGTATTTTGT CGGGGGCAGC	3720
GTCCGTGATG CTTTATTAAA TAAACCAATC CATGATGTCG ATATTGCGAC AAGTGCCTAT	3780
CCAGAAGAAA TTAAGCAAAT TTTTAAACGA ACCGTTGATG TGGGCATCGA ACATGGCACA	3840
GTCCTTGATG TGATGGAAGA CCAACAGTAT GAAGTAACAA CCTTTCGTAC CGAGTCAACT	3900
TATCAAGACT TTCGCCGACC AGATGAAGTC ACCTTTGTTC GTCCTTGAA AGAAGATTTA	3960
AAACGGCGAG ATTTTACGAT TAATGCCTTG GCGCTCGACA GCACAGGTGA AATCATTGAT	4020
TTATTTGATG GTATCGAAGA TTTAACTAAT CAAACCATTG GTGCAGTCGG TAATCCCaT	4080
GAGCGGTTTC ATGAAGATGC TCTGCGAATG ATGCGGGGcT TACGCTTTGC TAGCCAATTA	4140
GATTTTAAAA TTGAAGAAAA AACCTTAGCC gCTATCGCTG AATTTTCATCC TTTATTGGAA	4200
AAAATTTTCAG TAGAACGCAT TACAATTGAA TTTGTAAAAA TGCTTTTAGG GGTTAATCGA	4260
CAAGGTGGCT TAGCTCCGTT TATTGAAACA GAATGCTATC AATACTGTCC AAAATTACGA	4320
GAACAAGGGG CTGGACTTTT TCGTTTAATG GACTTGCCTG CTCGTCAAAT TGAAACAGAA	4380
GCAGAAGCGT GGACGTTACT TATCCAATCG TTGAATTTAC CAGAAGCAGA AATTCGCTCC	4440
TTCTTAAAGG CTTGGAAACT TTCAAATCAA TTAATCCAAA ACGTTTCACA ATTAGTACGC	4500
GGTTTGCGTT TTCGCTTGAG TAACGACTGG CAGCCAATGA TGCTTTATGA ATTAGGCGAA	4560
GAATCAGCAG TACTTGTTGA AAGGCTTTTA TATTATTACC AACAAGAGAG TCAAGTACAA	4620
GTAACAAAAG AGTTAGTTAA GGCATTACCC ATCCATCAA GACATGAATT GGCCATAACA	4680
GGAAAAGACC TTTTAGCCGT CTAGAAGAA ACACCAGGCA AATGGTTAGG GGAAGTATT	4740
GCTGAAATCG AGCAACACGT TGTTGAAGGT AGCTTAGAGA ATAAACAAGA AGTTCTGCTT	4800
TCGTTTGCCA AAAACAACG TTCAAAAGGA GAGAAAGCAT AATGCCAACA GAAATGTTTA	4860
CAAAAGAATT CTTGGTGAAA GAATCACAAA CGGCTArAAT GCTCGGTTCT GGTGACTTAG	4920
AAGTTTTAGG AACCCCTGCT TTAGCTGCCA TGATTGAACA AaCAGCCAAA GAaGCTGTGA	4980
AGGA'CCAwTT ATCAGTTGGG GAAaCaACTG TCGGAaCGGT GTTAGAATtG CGCCACTTAT	5040
TTCCATCCGC GGTAGGGGCG ACAATCGTTG TTACGATGAC CAGCATTGAA CAGACTGCAC	5100
ACAAAATACG CTATGAATTT GTTGCTTATG AAGGAGAACG ACAAATT	5147

## (2) INFORMATION FOR SEQ ID NO: 117:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 12022 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

AAGAATTTAG ACAACTATTA GCCGCAAAAG GCATTGAATT GACAGACCAA CAGATGACAC	60
---	----

AATTTGACCA ATATTTCCAT TTACTAGTAG AATGGAATGA AAAAATGAAT TTAACAGCCA	120
TTACAGAAGA AAAAGAAGTT TATTTGAAAC ATTTCTATGA TTCTGTTTCG TTGGCaTTCT	180
TTGAAGACTT TGCCAGTGAC AAAGCCATTT GTGATGTTGG CGCAGGCGCA GGATTTCCAA	240
GTATTCCTTT GAAAATTGTT TTTCTTCAT TAAATGTCAC AATTGTTGAC TCTTTAAATA	300
AACGAATTAC TTTCTAACA GAGTTGGTGA ATCAACTAGG TTTGGCTAAT GTTCTTTGT	360
ATCATGATCG TGCAGAAACA TTTGGTCAAA AAGCAGAATT TCGTGGCGCG TTTGATTACG	420
TTACAGCGCG AGCAGTTGCA CGATTAAATG TTTTAAGTGA ACTGTGTTTA CCCTTAGTTA	480
AAAAGAAGG CTATTTTTTA GCCTTAAAG CATCGAAAAG TGAAGAAGAA ATTAATGAAG	540
CAAAACCTGC AATCGCTACA TTAGGCGGTC AATTTCAAAA AGAAGTTGGT TTTACTTTGC	600
CGATTACGGC AGACGAACGC CACATTGTCG TGATTCAAAA GAAAAAAGAA ACCCCTAAAA	660
AATATCCCAG AAAACCTGGA TTACCAAATA AACAACCGAT TAAATAAGAT GGAGGCACAC	720
ACATGACACG AATTATTTCT GTAGCGAACC AAAAAGGCGG CGTTGGTAAA ACAACGACCA	780
CTGTAAATTT AGGCGCATGT TTAGCAAACCT TAGGCAAAAA AGTCTTACTG ATCGACATTG	840
ATGCACAAGG AAATGCTACC AGTGGCATGG GTGTTCTTAA ACCAGATGTC GCACACGATG	900
TGTATGATGT CCTTGTAAT GAAGAACCAA TCACTAGCGT GGTCAACAT ACCAGCCGTG	960
AAAATTTAGA TATTGTCCCT GCAACTATTC AATTAGCAGG CGCTGAAATT GAATTAACTT	1020
CAATGATGGC TCGAGAATCG CGCTTGAAAT TAGCAATTGA TGAAGTAAGA GACATGTATG	1080
ATTTCGTTTT AATTGATTGC CCGCCTTCTC TAGGGCATT TACAATCAAT GCTTTTACAG	1140
CAAGTGATTC GATTCTAATT CCTGTTCAAT GTGAATATTA CGCATTAGAA GGCTTGAGTC	1200
AATTGCTAAA TACGATTCGC TTGGTTCAAA AACATTTTAA CCCTGAATTG AAAATTGAAG	1260
GGGTCTTGTT AACCATGTAC GATGCTCGGA CCAATTTAGG TGCCGAAGTC GTTGAAGAAG	1320
TCCGTAAATA TTTCCGTGAA AAAGTCTATG ATACGATTAT TCCAAGAAAT GTCCGTTTAT	1380
CCGAAGCACC AAGTCATGGT TTACCAATCA TTGACTATGA TATTCGCTCA AAAGGTGCCG	1440
AAGTGTATCA AGCACTAGCA AAGGAAGTGT TGGAGAATGA ATAAAGGCAA AGGTTTAGGc	1500
AGAGGCATCG ATGCCTTGTT TCAAGATATC GCAaGCTAG AAGATGTCGA CGTTAAGaC	1560
GAACAAGTGA CAGAAATTCT ATTAAACGAA CTTCTGCCA ATCCTTACCA ACCGCGGAAA	1620
ACATTTGATG AAACCTCTTT ACAAGAACTT GCAAACTCGA TTCTACATTC CGGTGTGTTT	1680
CAACCGATTA TTGTTCTGTA ATCGGCAGTT AAAGGCTATG AAATTATTGC AGGTGAGCGT	1740
CGGTTCCGGG CTTCTAAATT AGCCGGCAAA GAAAAAATTC CCGCAATTAT TCGCGAATTT	1800
GATGAAGAAT CCATGATGCA AGTTGCTGTT TTGGAAAAC TACAAAGGGA AGATTTAAAT	1860
CCGCTAGAAG AAGCAGAAGC CTATGAAATG CTAATGAAAA ATTTAAATTT AACACAAGCG	1920
GAAGTCGCAG AACGTTTGGG CAAAAGTCGC CCGTATATTG CCAATTATTT ACGGTTGCTA	1980
ACTTTACCAG ATGCGGTCAA AGCAATGGTT CAAAAACAAA GCATGTCAAT GGGTCAAGCA	2040

CGGACGTTAT TAGGCTTGAA GAACAAAGAA CAGCTATTGC CATTAGCTAA TCGTTGTATT	2100
AAAGACAATT TAACGGTTCG ACAATTAGAA CAACTTGTCTG CTGAATTAAA CGAAACACAA	2160
GGCAAAAAGG GCAAAAAGC GAAAAAGCC ATCAAGGAAA AACCATTATA TATTTCGAGAA	2220
AGTGAAGACC GCTTAATGGA TAAATTTGGG ACAACCGTAG CGATTCAAGA AAAAGAAGGC	2280
AAAGGAAAAA TTGAAATTGA GTACTTGTCC TCTTCGGATT TGGCTAGAAT TTTAGATATT	2340
TTAGATATTC ATTTTGACGA GGAATAAAAA GCAAAGTGAG AAAGTAAGAC AGAAAGTCAC	2400
AAATCCTTTT GCTAGTTTCG GTAAAAATG GTATTGTAAT CATGGTATTT TTATATAAGT	2460
AAGAGAAGAA AGAAGGCTTG GAACATGTAT GATTTAGGCG ATATTGTCGA AATGAAAAAG	2520
CCGCATGCTT GCCAAGCAAA TCGTTGGCAA ATCATTTCGA TGGGCGCTGA TATCAAAATC	2580
AAGTGCACGA ATTGCGGGCA TATTGTCATG ATGCCGCGTC GTGATTTTAC TAAAAATTA	2640
AAAAAAGTAA TCGAGAAAAA AGCAGACTAA TGCTTTGATG CCTCTAATAG AAAAGAGAAT	2700
AGACTGAGAG AAAGAGTGGG AAACAAATGG CACTAACAGC TGAATTGTC GGCTTACCAA	2760
ACGTTGGTAA ATCGACCCTA TTTAACGCAA TTACAAAAGC AGGAGCAGAA GCTGCAAATT	2820
ATCCCTTTCG AACGATTGAT CCGAACGTCG GCATGGTTGA AGTACCAGAT GCCCGTTTAC	2880
AACGCTTGAC AGAAGTAGTA AAACCGAAAA AACAGTTCC CACAACGTTT GAATTTACAG	2940
ACATCGCCGG AATTGTTAAA GGTGCCAGTA AAGGAGAAGG GCTAGGAAAT CAGTTTTTAA	3000
GTCACATTCG TCAAGTAGAT GCAATCTGTC ACGTGGTTCG TTGTTTTGAT GACGAAAATA	3060
TCATGCGTGA ACAAATCGG GATGCGGATT TCGTTGATCC TTTAGCAGAC ATTGATACAA	3120
TTAATTTAGA GCTAATTTTA GCAGATTTAG ATTCAATCAA TAAGCGCTAT ACTCGCGTGG	3180
CTAAAATGGC GAAAGCAAAA GATAAAGAAG CGGTGAAGA ATTAGCTGTT TTAGACAAAA	3240
TTAAACAGT TTTAGAAGAA GGTATCTCTG CTCGTACGAT TGAATTTACG CCAGAAGAAG	3300
AGAAAATTGT TAAATCTCTT TTCTTATTAA CCACAAAGCC GGTCTTTTAT GTAGCAAATG	3360
TTTCAGAAGA TGAAGTGCG GATCCAGACA ACAACGAGTA TGTACAACAA GTTCGTAATT	3420
TTGCAACCAG TGAAAATGCG GAAGTAATCG TAGTTAGTGC GCGAGCAGAA GAAGAAATTG	3480
CTGAATTGGA TTCTGAAGAA GACAAGGCTG AATTTTTAGA AGCAATGGGC ATTGAACAAT	3540
CTGGCTTAGA TCAATTGATT CGTGCGGCTT ACGATTTGTT AGGCTTAGCT ACCTACTTTA	3600
CAGCTGGTGA ACAAGAAGTT CGTGCATGGA CGTTCCGTAA GGAATCAAG GCACCTCAAG	3660
CAGCTGGGAT TATCCACAGT GACTTTGAAC GAGGCTTCAT TCGTGCGGAA ACAGTTTCTT	3720
ATGAAGACCT AGATAAATAT GGCAATATGC AAGCTGCCAA AGAAGCGGGC CGTGTTCTGT	3780
TAGAAGGAAA AGACTACGTC GTTCAAGATG GCGATGTTAT GTTGTTCCGT TTTAACGTTT	3840
AAAAAATGT GCCCAACAAG CAATTTGTTT TGGGATAAGC CCACATATTT AAGGCCATTG	3900
AGCCGTTTAG AAAGAGGTTT CCTCTACATG AGGGTGAGGA GGATACAACT ATGGAAGCAG	3960
AAGCATTACG AGCAATCGTG GCTGAGAATC GTCAATTAGA ACAGAATTTG ACTAAAAGAA	4020



ATGAACAATA CATTTTTGAC TTAAAAAAT CATTAAAAGC GGCTAATTTA TCTGAAGAAG	4080
AGCTAGCATT GGCTCTACAC GGTATTTTGC CAGAATTAGT AGCCGGTCAA AAAACAGGTA	4140
AAACAGCCCG TCAGTTATTT GGCAGTGTTC CAGAACGGAC GGAAGCAATT TTAAATAAAC	4200
CAGCAGAAGT CAAAGAACCA GCGGGCTGGA TGATTTGGCT CGACAATACC TTACTATTAC	4260
TTGGTTTATT AACAAATCATG TTAGCGGCGA TGTCTCTATT TTCAAAAGGA ACCGCACAAC	4320
CTCTCGGATT AACCACATAT ATCTTAGGTG CGATGGCTGG GGGTTATGTT TTCTATCTAA	4380
TGCACAAATA TGTTTACCGT TTTGATCGTC AAGGTGGCGA CAAAAGCAA CGTCCTGGTT	4440
GGTTAAAAAC GACCGCCATT CTATTTGGTG GGATGTTCTT GTGGATTGCA GTTTTCGCTG	4500
GTTTCAGCAT GCTACCACCA GTGATCAATC CCATCTTAGA TCCGATGATT GCTTTAGTAA	4560
TCGGTGGCTT AGCCTTTGTA GCACGCTATT TCTTCAAGAA AAAATACAAT ATTCAAGGTA	4620
GTTTCATGAC ACGTCAATAA ACGGCGCGCG CTCAATTGTG AATGCGCTCG GcTGACGCAG	4680
AAACTGCAAC AAGCAAAGTC GCTCTCTCAA TAGAGCGTTT TAACAGTTAA AGCCTGGGGG	4740
ATTAGTCGAA AAAACTAGTG TCCCAGGCTT CTGTTTTAG AAGTTTGCTA AAGAAAGGAG	4800
CCTCTTATGG GACAACAAGT GAAGTCTTGG TGCCGTCAAC ACGACCGGCT ACTCATTTTT	4860
ATTAGTTTTT TCTTATTAAG TGGACTTTCT CTCTATTTTG TCTATTTTCA AGAGAATTTT	4920
TTACGAGCAA GTGATTTCCG CTTTCATCAA AATCGGGTGG AAGGTCTCGC ATTAGCAATT	4980
AAAAACAACG ACTGGTTTCC TAAAAATCAAT TATTTTTTCT TGGGAGGCTA TGGCTATGCC	5040
TCTAGTTTGT TTTATCCAGA TGCGTATCTT TATTTACCTG CGCTTTTGCG TGTTCTGGGG	5100
ATTTCTTTG TCGCTAGTAT GCGGATTTTT GTTTTTGCGG TAAACTTGGC AACATTTAGT	5160
CTTACTTATT ATGCAGGCCG ATTAATGGCG CTCTCTAAAA AAAGAAGCTA TCTCTTTGCC	5220
ATTTTGTATG GCTTATCTAT TTACCGCATG CAAGACTTGT TTAATCGCCA AGCGTTAGGA	5280
GAATTTTTGG CGCTAAGTTT TTTCCGTTG GTCTTAGCGA GCCTGTTTTT ATTACGAAAG	5340
GGAACCACGA AATGTTGGCC GTTGTTAACG CTGGCGATGA CTGGGATTGG GTTGGCTCAC	5400
TTTATTTCTA TTGAAATGGT TTCTATTTGG ATTGGGTAT ACATTCTGTT TTATTGGCAG	5460
CAGTTTTTTA AAAAGGAGGT CCTGTTGGCA TTAGCGAAGG CTGCAGGTTT GACGCTGCTT	5520
TGGCTGGCCT TTTATCTTTT ACCAGTTGCT GAACAAATGA AAAATCAAGT GTTTAAAGTT	5580
ACGTCCAATC CACTCACTTA TATTTCCGAA AGAAGTTATC CTATTGTTTC TCTTTTCATT	5640
AATAGTTTAA AGAGCAGTGT TTTTCACGCG AAAACAGCGA ATCTAGGAAC CCTTTTGTTT	5700
GTGGGGCTTG TTGTAGCAGT GGTTCGCTT GCCAGCAAAA AAATTCAGAA CAAACGGTTT	5760
ATTGGTTTGA CGCTGGTACT CTTATTGATG GTTACCACGC TTTTTCCTTG GCACTGGCTG	5820
AACCACACGC CTTTAAATAC AATTCAGTTT CCTTGGCGGT TTTTAGGGAT TCTTTCTGTA	5880
ATGCTCGCTT TTTTCATTGC ACAAGACGAA TGGGGCGTTT TCCGTAAATC TTGGACCGTG	5940
GCTTTGTTGG TTTTCCTTGC CATCAGCAAT TTGGGCATTT ATCAATATCA AAGTATTCAG	6000

TCACAACAAG GCCGCCTGTT AACAAAAGCT GAGTATGAGC AGCCAGCGCC TTTTATATATC	6060
GGTGCGGGCC ATGAATATTT GCCAGATGAA ATCAATTATC AGGAATTATT GAAACAGAAA	6120
AAACGACCAC TTGATTATTC GGCAGAACAA GTAACCATTA CAAATGTTAG AaTGCCTTAT	6180
GGGAAAATTT CATTGACTA CCAAGTGGTG AtCAATCAGC GAAGGTGACC GTTCCTTTTA	6240
TTTATTACTT AGGCTATCAA GCCACGATTC AAATGAAAA CCAAACGGGC GCTAAAAAA	6300
TGAGCCTAAC CAATCAAGGA GGCTTAACCG CTCTTTCTTT ATCTGGTACA GGACATGTCTG	6360
ACATCCGTTA CCAAAGAACA AAAGTGCAAA AAATAGGGAC AATGATTACC TTACTTTCTG	6420
TTGGTGGTTT TGGATTTAGC CGATTTTTAC AACAAAAAA GAAACACAAG ATAAAAAGAAC	6480
AACGATGAAA GTAGGACCTT AATTAGGAAG AGTTGAATAA AGGACAAACC GATTTTGGCG	6540
TTTgTCC <sup>1</sup> TTT ATTCTCATT GATTATTTAA ACAAATCTAG TTGACTTCCT TAGTATTATC	6600
TGTTATTTTC ATATATAAAA TTAAGTCGAG GAGTGGACAA AAAAATGTCA AaY <sup>1</sup> TGGGAAA	6660
CAAAATTTGC GAAAAAGGT CTTACATTTG ATGATGTGCT ATTAATCCCT GCAGAAAGCC	6720
ATGTCTTACC AAATGATGTA GACATGAGTG TACAATTAGC TAAAAATATC AAGTTAAATA	6780
TTCCATTGAT GAGTGCCAGT ATGGATACTG TTACAGATAG TAATATGGCA ATTGCGATGG	6840
CTCGTCAAGG TGGACTAGGT GTCGTACATA AAAATATGAC TGTCGCCCAA CAAGCAGATG	6900
AAGTACGTAA AGTCAAACGT TCTGAAAGCG GCGTTATCAT TGATCCATTC TTTCTAACAC	6960
CAACAACTT AGTAGCGGAT GCAGAAGAAT TAATGAGTCG TTACCGTATT AGCGGTGTGC	7020
CAATTGTAGA AACAAATGGAA AACCGTAAAT TAGTCGGAAT TATTACAAAC CGTGATATGC	7080
GTTTTGTAAC GGA <sup>1</sup> CTACCAA ATCAAATCG AAGAAGTCAT GACAAAAGAT CATTTAGTCA	7140
CTGCGCCAGT TGGCACATCT TTA <sup>1</sup> AAAGATG CAGAAAAAAT CTTACAAAA CATAAAATCG	7200
AAAAATTACC AATCGTTGAT GAAGCAGGTC GCTTAAGT <sup>1</sup> GG CTTAATTACA ATTAAAGACA	7260
TTGAAAAGT GATTGAATTC CCA <sup>1</sup> AATGCTG CCAAGATGA ACACGGTCGT TTA <sup>1</sup> CTGTGG	7320
CGGCAGCAGT CGGTGTCACT AGTGATACGT TCGAACGTGC GGAAGCATT <sup>1</sup> A TTAGAAGCAG	7380
GTGCGGATGC GATTGTCA <sup>1</sup> TC GATACTGCAC ATGGCCATAG TGCCGGCGTT ATT <sup>1</sup> CGAAAA	7440
TTAAAGAAAT TCGTGAAACG TTCCCTGAAG CAACATTAAT TGCTGGTAAT GTTGCCACTG	7500
CTGAAGCAAC CAAAGCATTG TATGATGTCG GCGTTGACGT TGTCAAAGTT GGTATTGGAC	7560
CTGGTTCAAT TTGTACAACA CGTGTCGTTG CTGGTGTGG TGTTCTCAA TTAACAGCGA	7620
TTTATGATGC TGCTTCTGTA GCTAGAGAAT ATGGTAAAGC AATTATCGCT GATGGCGGTA	7680
TTAAATATTC TGGTGATATC GTGAAAGCCT TAGCTGCTGG TGGACATGCT GTAATGCTAG	7740
GAAGTATGTT AGCAGGTACA GATGAATCTC CAGGCGAATT TGAAATTTAT CAAGGCCGTC	7800
GTTTCAAAC TTACCGTGGC ATGGGCTCAT TAGGTGCAAT GGAAAAAGGA TCAAGTGATC	7860
GTTACTTCCA AGGAAGTGTC AATGAAGCAA ACAAATTAGT GCCAGAAGGA ATCGAAGGAC	7920
GTGTTGCTTA CAAAGGAAGC GTTCTGATA TTGTCTTCCA ATTAATCGGT GGCTTGAAAT	7980

CAGGGATGGG TTACGTCGGT GCAGCAGACC TGAAAGCGTT GCGTGAAGAA GCACAATTTG	8040
TTCAAATGAG TGGTAACGGA TTGAAAGAAT CACATCCACA TGACGTTCAA ATCACAAAAG	8100
AAGCACCAAA CTATTCTGTT CAATAAAATA AAAGCTGATT TTTTGTAAAT TTAAAAGAGG	8160
AAGACGTCAT TTCGACGTCT TCCTCTTTTT CTGTAGAATG AAGGATTTGC TTCTGTTTCC	8220
GCCGTTTATC AATTTTTAnG CCTGTAACAA AAaTCCAAAG TGATTTTTGT TACAGGcCC	8280
TTTTCCCTATA GAATAAAAGC CGTTATCTGA TATCAGACAA CGACCTTCAT CTTCCCTTTA	8340
TTTAATAACT GTTAGGTTAC CCATGTAAGG AACTAGGACT TCTGGTACAG TTACAGAGCC	8400
ATCTTCGTTT TGGTAGTTTT CTAAAATAGC AGCGACTGTC CGTCCACGG CTAAACCAGA	8460
ACCATTGAGT GTATGCGCAT ATTGGACTTT ATCATTTTCA TCACGGTAAC GAATCATCGC	8520
ACGGCGCGCT TGGAAATCTT CACAGTTTGA GCATGAACTG ATTCACGGT ACGTCTCTTG	8580
AGCAGGAATC CAAACTTCCA AGTCATAAGT TTTCGCAGCT GAGAAGCCCA TGTCACCTGT	8640
TGAAAGAGCC ATGACACGGT AAGGTAATCC CAGTTTTTGC AGAATTTCTT CGGCATTATT	8700
CGTCATTTTT TCTAATTCCT CATAAGAATG CTCTGCATCA CTAAATTTAA CCATTTCGAC	8760
TTTATTAAAT TGGTGTAAC GAATTAATCC TCGTGTGTCA CGACCAGCGC TACCAGCTTC	8820
AGAACGGAAA GAAGGGCTCA aGGCCGTAAa GTAAATCGGT AAaTCCTTGC CATCTAAaT	8880
TTTATTGTTG TAGTAATTGG TTAAAGGGAC TTCCGCAGTT GGAATTAACG TTAAATCCGT	8940
ATCTTGTAAT TGGAAGACAT CTTCTTTAA TTTAGGAAAT TGGCCAGTCC CGAACATGGC	9000
GGTGTCAATTA ACGATATAAG GCGTAATCAT TTCTGTATAG CCATGTTTCA AAACATGTAA	9060
ATCTAACATG AAGTTGTATA ACGCACGTTT TAAGCGTGCA CCTAAGCCTT TGTAATAAAC	9120
AAAACGGCTA CCAGCTACTT TGGCACCACG TTCAAAATCT AGAATGCCTA AGTTTTCAGC	9180
GACTTCCCAG TGTGGTTTTG GTTCAAAAGC AAAAGTTCTT GGTTCCTCC AACGACGTAC	9240
TTCTACATTT TCTTCTTCAC CAGCACCAAC AGGCACAGAA TCATCTGGCA AATTCGGTAA	9300
AGTAGTTGTG ATCCCACGTA GCTCTTCGTC AATAGCGTTA ATTCTGCAT CTAATGCTTT	9360
AATATTGCCG CCGACTTCTT TCATTTCCGC GATTTTGGCC GTTGCACTCT CTTTGTTCG	9420
TTTTAATTGA GCAATTTTCA CAGACACATC ATTCCGGTAT TTTTTCATTT CTTGACTTT	9480
GACTAATAAA TCACGGCGAC TTTCATCTAA ACGCAGAAAT TCCACTAAGA TTTCTCTTT	9540
CACGCCACGG GTTTGTAGCT TTGCTTTTAC TTCGTCAAAA TTTTGACGCA TCATTTTAC	9600
ATCTAACATA TTTTTTCCTC CTTAAATTA GTACATAATC AATAATAACT GTTGTAGAAA	9660
ACAACAAAAA AGCCATCTCA CCCCAAAAT AAATATTCTT GGGACGAAAT GACTGGAAT	9720
CTCGCGGTAC CACCCAAGTT CAGTAATGTT ACTGCCCTTG ATTTGCAGAT AACGGCTGCA	9780
GACCGTTCCA ACTTATTTTCG CTGGACACCT GGAAGAGACG GACTTCTTTA TAGATGAATG	9840
TTTGCTCGCA CCCCCGAAA CTCTCTGGAA TTCAGTGCCA TAAATACTTT TTCTCTTAGT	9900
TTTCTCTTTT CTATAGAATA CCGAAAAAGT TCAGGGATGT CAATTAATCT TCTGATTCT	9960

TGTTGGTACG	CTGTAAGGGT	AGATGAATAA	TAAAGGAAGT	CCACTTTTGA	TCAGACTTCG	10020
CATAAATATA	TCCACCATGT	AAAGCGACAA	TACTTTGTGC	AATAGCCAGA	CCTAAACCAG	10080
TACCCCCCGT	TTCTTGTGAT	CGGGATTCTT	CGACACGATA	GAAGCGGTCA	AACAACGTAT	10140
CTAAAGAATG	TTTAGGGATA	GCTGGTCCAT	CATTGCGAAC	AGCAATAATT	GCTTCGGTGC	10200
CGACTTTGTC	TACTTCCATT	ACAATATGAT	GGCCACCTTT	GCCATATTTT	AAAGCGTTTG	10260
AAAGAAGGTT	GTTAAAGACT	CGGACTAGTT	TTCCGTATC	GCCTTCCATC	ATTAAAGAAG	10320
CTGGATTGGC	TTTTACTTGG	ATTTGCATAT	TAATCTTTTT	CGCTTCCAAT	TCAAAATCGG	10380
CAGCAAGCTG	TTCAATTAAT	TGTGCCATAT	CAAAAAGTCG	CGTATGAATC	GGTACACTTG	10440
GTTGACGGAC	TTTTGTATAT	TCAAAAAGAT	CATCAACTAA	TAATTTTATT	TGTTTCGCTT	10500
TGACGTAAGC	GGTATGCGTA	TATTTAGTA	AATCTTCTTG	GCTATGAAAT	TGACGATCTT	10560
CAATCAATCC	CAGATAACCA	ATAATAGAAG	TTAACGGTGT	CCGAATATCA	TGACTGACAT	10620
TCGTAATTAA	TTCATCTTTT	GATTTTTTCA	TTCCCGTTC	ATCTTCAATG	GCTGCGACAG	10680
TACTATCCAC	CAGACCGTTA	ATACTCGTCA	CGACTTTTGC	TAAATCACCA	CTTAATTCAA	10740
ACGGAATGCG	GTGATTGTAA	TTACCATCAG	CGATGTAATG	TAACACTACTG	ATAATATGAC	10800
GCAATTGCAT	TTGATGGTAG	CGACGAATTA	AGCGCCAATA	GAGAACAGCT	GCATCTACTA	10860
AAAAGAAAAT	GGGTAAGACA	AATTTATTCC	AGCTCCAAAA	CAAATCAGTA	TTTAATTCTT	10920
TAGCGACCGC	ATCCTTAGTG	GCCCAGATGA	CACCAGCTAA	GGGCGAGTTT	TCACTAAGTA	10980
GTTGTCCTAA	AAGGAACATA	ATTGAGACAT	TTAATAGTAG	TAATAAAACA	ATTGTCACAA	11040
TTGCTTCTGC	TAATAACTCG	CTGATTTCTT	TAGAAGTTAA	GGTAATTCGT	TTTTGTTTTT	11100
CAATCTTTTT	AGGTTTAACG	AGCATCGATC	TTATATCCAA	CTCCCCACAC	TGTTTGAATG	11160
ACTTTTTCTC	CACCAGTTGC	TTCTTCAATC	TTATCTCTTA	AATGACTGAC	GTGAACCATC	11220
ACGGTTTTAG	CAGAAACAAT	ACTTTCTTGT	TGCCACACAC	GCTCAAAAAT	TTCATCCGCA	11280
CTAAACACAC	GATTTGGATG	ACTAGCTAAT	AGATAGAGGA	TACCAACTC	AAGAGCGGTT	11340
AATTGAATTT	CTTTTCCTTC	TATTGTTTTG	ACTTCGTGAG	AGTCCTTATT	AATCATTAAT	11400
GGGCCGACTT	CTAATTCATC	GGGTTCTTCT	TTGGTTAATT	GCATTTGGCT	TCTACGTAAA	11460
ATAGACTTTA	CCCAGCCAT	TACTTCTAAA	GGATTAAAAG	GCTTGGTGAC	ATAATCGTCT	11520
GCACCAGCAA	CTAAGCCTTT	GATTTTATCC	ATATCGGTTG	TTTTTGCTGT	TAACATAATA	11580
ATTGGAATTT	GCGATTCTTT	TCGAAGTTCT	TTCCTACTTT	CCATTCCATC	CATGATTGGC	11640
ATCATAATGT	CCAAGATTAA	TAAATCGATA	TCTGGTGTTG	TGTGCAATTT	GGATAACGCT	11700
TCTTTTCCAT	CATAAGCTTT	CACGACTTCG	TATCCTTCAT	TATGAATATA	AATACTCAAT	11760
AGTTCTACAA	TTTCTTTATC	ATCATCTGCA	ACTAAAATTT	TCATAAGCGT	TCATTTCCCTC	11820
CATATATAAA	AAATCTCTTT	TCCATATTTT	AACAAATAAT	TGCTAAAAAG	GACAAAATCC	11880
TTGATATATT	ACGATTAATT	TAGAAGATTT	TGCTAAAAGT	TAATAAACTA	TAGCCAAAAA	11940

AGGATGAAAG GGGGTAAAAA ACGAACGAAA ATACCTATAC CAATCATAAT AATCGGTATT 12000  
 GAAAAAAAGT TTAaaaaaaAG TG 12022

## (2) INFORMATION FOR SEQ ID NO: 118:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 556 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

TAAATAATC GCAATGACTG CATCTGTAAA TGCTTCTACA CGTGTTCG GCATCATTTT 60  
 CCCCTACTTT CATAAAATT AAATACTTA CCCAAAGTAA GTATAACGCA TTTCTGTAAA 120  
 AATTCAATTA TTTATGAAGT CTCTCGGTTG TTGAAGTTCC TAAAAACGAG TGCAAGAGCC 180  
 CCACTTAAAA GTAAGCCTCT TGCACTCGTT ATAGAACAAA TCAGTTATTG GAAGACGCCT 240  
 GCCTGATAAG GTGCTAACTG CATCCGTTTTC TTCACCGTAA AATCTTGATT CCCTAATAAT 300  
 TGTGCGCACT TTCTGCCTAG TAATTTATTC GGCAAGTCAA ATTGGCACTT TTCTCCACTT 360  
 AAATTGACAA TCACTAACGC TGTGCGCTCA GTTGTTCCTT TAAAAAAGC AAAAACAGAC 420  
 GGATGATTGG GcGCTAATAG TTCAAAGTGT CCACTGTGTA AGACCGGATT GTCTTCCGC 480  
 AACGCAATTA ACTTTTtATA ATAATcAATA CGGATgCGCA TcCGTTtCTT GGGCAGCCAC 540  
 ATTTATnnCT TGAAAT 556

## (2) INFORMATION FOR SEQ ID NO: 119:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5998 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

TGTTAAAAAA ACCATTTTTTA CTATTTTTTTT CACTTTTAGG AGCTATTTTTT ATTTTAGCTA 60  
 GTTGTGGCAT AGGAAAaGaT GCTGTCACAG ATACTAAGTA CAAAGTTAGT TTGCAGCAAG 120  
 CTGCTGAAAT CTATGAAAAA GAAGCTGGCA ACAGCAAACC ATTAGTAAAT GTCCAATTTG 180  
 ATACAGAACC AGCAAGTGAC TACAGCTATA TCTTTACTAA CGATACAGAA ACACTTTACG 240  
 TGAATCCTGA AACAGGAAAA GTCACCAAAA ATACTGAAGC AAATCAACTT GGCGAAAACG 300  
 AGACAGCCTT TTCAGCTGCT GAAGTCAAAG AATTAGGCGC TGTTAACGAC GTTTTAGCCA 360  
 AAGCAAAAAA AGAAGTTGGA GGACTTTCTC CACGTATTTT GACTTGGAAG TTAACCAAAA 420  
 ATAACAATAA ACTTGTTTAT ACAGTAGATG TTAACGAC TACGGCAGAT GAAAAAGTTA 480  
 CCTTAGATGC CAATAAATAA GCACAAGCGT ATAACCTATC ACGAAATAAC CAAATCTTAA 540  
 AATACTCTTT GGGACAGAAA CTACGCATTA GGAAACCTTA ACATTTTCCG AAGCCGATCG 600

TTCTGTCCCA	ATTTCAATTCT	TTTAAGTAAT	AATAATTGTT	CGTTACCCCT	GCTACCGTTA	660
TACTTAGTTT	AAAAGAAAAG	ATGGGAGTAT	CAATTCATGG	TTTATACAGT	TGATTTTAAA	720
GAAGTAGCCA	CTATTGGTTT	AGAAAGTTCA	CCAGTGGCAC	CTGCCCTTGC	TGGACTAAGA	780
GCGAATGAAG	CTCGCTATTT	TTGGAATAAA	TACAAACATC	GCTTTACGAC	CGTCCCTGCC	840
TCTGAAGCTC	CAGAAACATT	GGCCTGGATT	GAAAAAATTT	TATTAGAACG	AGAATTACAT	900
TTCCCTTACA	AGCCACTTGA	AATTTCTTCT	TTCACTTTAG	ATGGCATAAA	AATGGCCTAT	960
GTATTTTATG	AAAACGGCTT	AGCAGTTAAC	GTGATGTACA	CATTAGCTGA	GGAAGGCAAA	1020
CGCGCCGTTG	GGTTTAAACT	TTCAGATGGT	ATGGCAATCC	CTGCAGAATT	TGAAGGAAAA	1080
TTCAAATTTG	CTCGTCAAAA	ATCAAAATTA	GCTGGAACCA	TTGCGGGCTC	TTATTTTGTC	1140
ATTAAAGGAA	ECTATTAAAA	AAATGCTTTT	GCCCTTATTA	TTAAAGGCAA	AAGCATTTTT	1200
TTATTTTAAT	TCTTCAATAT	AGTTAAATAC	TTGTTGTCCA	GCAATACCGC	CTTCACCGAC	1260
CGCTGTCGTG	ATTTGACGCA	ATGTTTTTTC	ACGGACATCA	CCAATGGCAT	AAACACCAGG	1320
AATTTTGTG	CTCATTTCTT	GGTCTGTTTC	AATCCAACCA	GCCTCATTCG	TAATACCAGC	1380
TTTTTTAAAT	GGTCTGTGTA	ATGGATCTAG	CCCCACGTAA	ATGAACACGC	CGTTTGCTTC	1440
AATTTAGAA	ACTTCATCTG	TTTTGACATT	CCGTGCTTTC	ACACCTGTCA	CGACCATTTT	1500
ATTGCCAACA	ATTTCTTCGA	CAACTGTATC	CCAAACAAAG	GAAATCTTTT	CATTTGCAAA	1560
GGCACGATCT	TGAATGATTT	TTTGCGCACG	CAATTCATCA	CGACGGTGAA	CAATCACTAC	1620
TTCAGAAGCA	AATTGCGTTA	AATAGATTGC	TTCTTCAACT	GCAGAATCGC	CGCCGCCAAC	1680
GACTACTAAG	CGTTTATTTT	TAAAGAAAGC	ACCGTCACAA	ACAGCACAA	AAGAAACACC	1740
ACGGCCAGCA	AATTCTTCTT	CGCCTTTTAC	ACCCAATTTT	CGGTGTTTAC	AGCCTGTTGC	1800
AATAATTACC	GCTTTAGCTT	CATAAGATTT	GTCATCACAA	ATCACTTCTT	TATAGGATCC	1860
GTGGTCTTCA	ATGCCCATCA	CAATACCATA	CGCATTTTCC	GTACCAAATT	TTTCTACATT	1920
TTCATACATT	TTATAGGCCA	ATTCTGGACC	CATAATAGAA	TCAAAGCCTG	GGTAATTCTC	1980
TACCTCAGCA	GTGTTGTTCA	TTTGCCCAAC	AGGCGCACCG	CGTTCAATCA	TTAATACAGA	2040
AAGATTGGAC	CGTGATGCAT	AAAGTGCTGC	CGTCATCCCA	GCAGGACCGG	CACCAATGAT	2100
AATTACATCA	TACATTTTAT	TTCCCCCAAA	ATTTTAAGTA	AGTTTACTTT	ACTCCTTAAA	2160
CAATTTCTTG	TCTACTTTTC	TGCCCTTTT	TGACATTCGG	AGTAGGTGTG	TCTATTACGT	2220
AAGTTATTGT	ACCACGGTTT	TGTGTTTTCT	AGGAGGGGGA	AAAGAAAAGT	TTAAAAATAA	2280
CGATCACAA	GAATCATTGA	TTTTTAGAAG	CCGTTGTACT	GCCGCATCAG	AGAATTCTTC	2340
TTTATCATTT	TTTATGCATT	CTTGTTGATT	TTGCCCTTGT	GGAACAAAC	GAAATACTGT	2400
AAATGATGCA	GATAACGTAT	CAGGATTACT	AATCCGATGA	AGACTTTTAT	GCATCCGAAT	2460
AAGTTCTCCT	TCTGCACAAA	TTCTTGAAAT	AATTTGCTGA	TTTTCTAGCT	CTTCTATTTT	2520
CACTTTTCCC	GACTGCACAA	GTAAAACTTC	TTCGATTTTC	TGATGCTTAT	GCCAACTTGG	2580

CACAGTTTTT	CCTGGCAAGC	TATTTATATG	TATTTCAAAC	TCTGGAAATA	GAAAATAATT	2640
TACTTCTGTT	CCATCTGGCT	TTTTACGAAA	GATGACCTCG	TCTCTTGTTA	AAATTTCTAT	2700
ATATTTCTGT	AACATGGCAG	GCTCCTTTTA	ACTTTTTTTA	TAGCCGAAAG	CTATTCTTTC	2760
AAACTAAGTT	TGAAGCTACT	TTCTGAAAAT	TCGAGTTCCT	TTAAATATAC	TTTTATTACC	2820
ATTTGAAAAT	CACTTTTTTT	AAAACAATTG	ACTGTTCACT	ATCAATCTTC	ACTAATCCAT	2880
CTGTTCACCA	GAAAAAATAT	AGTATTTCCG	ATTATTCACG	AACTTATTAT	TCAACTTCAA	2940
AGGAAAATCA	GTCAACATAC	AGCCACATA	TCTAACATCA	GAAACACTTC	TACTTTTCTC	3000
AAGTACTGCT	ATGGTATACA	TTGACCTTAC	TAGTAACATT	ATAGCATCAA	GTTTTAATCA	3060
TATGTAGATT	TAAAAAAAAT	GaTAGcACCC	TAACAGaTGT	TTTCGTTGCT	CATTTAATTT	3120
ACATGATTAT	TTTTTAGATG	CTAATTAAAT	GTTTGTCACT	AACCTGAATT	TTTGATATTT	3180
TTATCCAGAT	TACTTATGTC	GGTTTATCAG	TTGTGCTAAC	AAAAAATGAT	TGTTTTTAAA	3240
GATAATTTAA	TTCATAATAT	ATATAGAATG	AAAAAAGGA	GGTTTTATTG	TTGAAAGTTG	3300
CTATTTGTGA	TGACAATCCA	ACGTTaACCG	AAAAAATAAA	CACGATGCTA	TTTAATTATA	3360
ACCCAAATAT	ATTTGAAACC	TATACCTACT	ACAATCCTCT	CAAGTTAGTC	AATCAACTTG	3420
ACCAAGAAAA	TTTTGACTTT	TTTATTTTAG	ATATAGAAAT	GAGTGAAATG	AGCGGAATAG	3480
ACTTGGCAAA	AAAAATCCGA	GAACGTGATA	TTTTTATCACC	TATCGTTTTT	TTGACGAGTT	3540
ACAAAGAATA	CATGGAAGAA	GTCTTCCAAG	TACAACTTTT	TGATTATTTA	TTGAAACCGC	3600
tACGGAAGAT	CGGTTGCACC	AAGTTTTGGA	AAAAC TGAGA	CAACAGCTAG	AAAAAAAAG	3660
AAATTACTTT	GTATTTTCTA	GTAACAAAGT	TACTTATAAA	ATACCAACCA	AAGATATCAT	3720
TTATTTCGAA	AAAGATAAAA	GACAAGTATT	GATT CATA CA	GTTGGAGAAa	TCTACAAACC	3780
CtATATGTCA	ACGAATCAGA	TAAACGAACA	ACTTGATACG	AATTTTGTGC	AAGTCCACTC	3840
TTCTTTCATA	ATAAATTGCG	CTTATATAAA	AGAGTTAGGA	AAAAATTTCC	TATTAATGGA	3900
TTCTAAAGAA	AAATGTATTG	AAATACCAAT	AAGTAGGCGC	TTCAAAGCCG	CCGCTCATAA	3960
AAGTATAGTT	ATGTCCATGA	GAGGAAAGAT	ATAATGCTTG	ATATAACATT	ACGAAGCCTA	4020
AGTTTGTTAG	TTTATTTATT	TCTTATCTCT	TATTTTTTTT	CTAATTGGTC	AACAGCTATT	4080
AAAAATCATA	GAGAATTTTA	TAATGTCATC	TGGATTGTTA	TAATCATATC	AGCCATGGTT	4140
ACGATTATAT	TATCCAATGC	TTACATACCG	AACATTTATT	TACCGTATCT	TAAC TTATTA	4200
CTAAGTTTCA	TTTTAACTTA	CAC TATTGCT	ATGCTACACT	CGGTCCCTTA	TCTCGACTCG	4260
GCCGTATGGA	CAATGACACT	TATATCAATA	AATTTAATCT	GTGAAGTATT	GTCGTTACAC	4320
TTTACGAAAA	TAATTTTAAA	TGCAGAATTA	ACATCCTATG	G TAGCCCTAC	TTTTTTCATA	4380
ACTTCAATAA	CAATTACAAC	TTTGATTGGG	G TAGCCTGGA	TAGTCATACT	TAAATTTTCA	4440
ATAGTTAAAG	AAGCAAAGTC	TGAATTATCG	ACAAACTTAT	CATTTATTGT	GATTTTATTA	4500
CCTATTCCCA	TACTATCAAT	AATCATTTTA	TTTGGGCTTT	TAATAGGAAA	CAACAATGAC	4560

AAAATTTCTG AGATTACTAT TACGATTAGT GTCATATTTT TAAACATCTG TGTAATTTTT	4620
CTTTATAAAA TCACTATAGA ATATCAAAAA AATCAATATA ATCTTACTTT GAGAAAAAAA	4680
AATATCGAAG TGGAGTATAG AATTTTAAAT GAAATAAAGA GAAACAGAAC TAACGTTTTA	4740
AAATTTAAAC ATGATTTGAA GAATCAATAC TTAACAATCC TCGGACTTAT CGAAAATGAA	4800
GAAGTTAATG AAGCTATCGA TTATATTAAG AGTAGTTTGT ATATTTTAGA ACCACCAACA	4860
AAAACTTATG CTGCCGATGG AGTATTAAAC TACCTATTGA ATGAAAAGCT TGCTGAAGCA	4920
AGAAAGAATC AAATTAATGT TGATCATCAG ATATTTGTCT CTAAAAATAT AAAAAATTAAT	4980
AATGACGTTT TAACTATTGT AATTGGAAAT ATAATTGATA ACGCCATACA AGCATCGAAG	5040
AGAATCAAGC CGATTGACAG ATATGTCAAT ATAATTATCA AGCAAGTCAA TAATGATTTA	5100
TTTATTGAAG TATCtAATAA CTATAATTCT GaAGAnATTT TtCaAGAAA ACaTAGaAAA	5160
AATaAGGtT TAGGaATGaA AAACaTTGaT GATCTCTTAC AACAAATTAGG TGGTATTCAA	5220
CGTCACTGGA CGAAGAACG TAGATATTTT GTAACATAG TTATATTTAA TGTATACAAA	5280
GGAAACAAAT GCGGTGATAA CTAATCAAGT AAATCACCGC ATTTTTTAGT ATACTTCCAT	5340
GTCGCATAAT TTATCAGAAA ATTTCAAAAA TAACCAAACA AGCATCAATC CCATAATCGT	5400
TGAAAATAGG AACGTTATAG TTGGAAAATT CATAAGACCC ATAGCAGAAA AATTCGAATA	5460
TATTAGCATT AGTACCACTG ACGTTATTAT TGAAGCAATT TCCGATTGTT TCCAAATTGA	5520
AATTATTTCT GACAATATCA AAATTACGAA TAGTAGACTA CAACTTGTA CTATGGTAAT	5580
AAGACCAAGG TAAATTTCAA AAAATAAATT TGAACATGA AACTTCAAAA TTAAGTTAAG	5640
CAAAACCTCA GTCACAAATG AAATAAGAAG CGCACTTCCA AAGCTAGATA AGATTATCAC	5700
GTAAAAGAA CTTGTTTTTT TTACGAGAAG CTCTTTTCGA TCAATCGGAA AAAGAAACGT	5760
TCTATTCCGG TGATTGCCAA CATAAAATCG TACAAGCTCT TTCCTTAGTA ATACAGCTCC	5820
ATAAATTGAT ATGCCGCACA TAGAAACAGT GTTCGATAAT GCTATAATTC CTTGAAATGA	5880
CGAGATCGTA TCAGCACTTT CATCAAATAT TCCCACAAAT GAGAAAAmCA TAAAAAATAA	5940
CATTGGATT ACCAGTGTA TTAAGAAAAC TTAAACCAA CATTnTAATG GnnCTAAT	5998

## (2) INFORMATION FOR SEQ ID NO: 120:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1641 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

AAATCCGTAG CAACATTTAT GGAGACAATT ACAGAAAAAT GCGGTGAAGC ACATCAAGAT	60
TGGGCAAAAA ACTTTCAGGG CGGnCTTTGC CAATACTTTA CTAACAACAG TGAAACGTGC	120
AAGAAGATGG TACTACCTTT TTATTAACAG GCGATATTCC AGCAATGTGG TTGAGAGACT	180



CAACAGCACA AGTTCGTCCT TATTTAGTGA TTGCAAAAGA AGACGAAGAC CTTGCTCAAA	240
TGATTGCTGG TTTGGTTAAA CGCCaTTTCG tTATATTTGC ATTGATCCTT ATGCCAATGC	300
GTTTAATGAA ACAGATAATC ATGCAGGCCA TCAAAGTATG AAGACAGAAA TGAATGGCTG	360
GATTTGGGAA CGGAAATATG AAATTGATTC ATTATGTTAC CCGGTCCAAT TAGCCTATTT	420
ACTTTATAAA AATACTGGGA TGACCGAACA ATTTAATTCA GATTTTGTGG AAGGCGTCAA	480
AAAAATTCTC AACGTTTTTA CAACAGAACA AGATCACGCA CAGTCACCTT ATTTATTTGA	540
ACGAGATACG TGGCGTCAAG AAGATACGTT GGTGGAAGCA GGGAAAGGAA CACCAGTTGG	600
CAAAACAGGG ATGACATGGT CAGGCTTCCG TCCTAGTGA TATGCCTGCC AATATGGTTA	660
CTTAGTTTCT TCTAATATGT TTGCAGTCGT AATTTTAGGC TATATTCAAG AAATTTTCTC	720
TGATGTTTTA GAAGATGCTA CGATTGGGGC GATTGCTAAA AAATTACAAG AAGAGATTGA	780
AGAAGGCATC CAAACATTG GACGAACAAA AAATCAAAAT AACGAAACGA TTTACGCGTA	840
TGAAGTGGaT GGCTTAGGCA ATGCAAGCGT GATGGAaTGAC AGTAATGTTT CCAATTTAGT	900
AGCAGCCCCCT TATCTGGGTT ATTGTTCAAC AGAGGATGAA CAATATTTAA CAACGCGTCA	960
AACCTTATTA AGTAAAGAAA ATCCTTACTT CTATGAAGGC AAATACGCGA AAGGGATTGG	1020
TAGCTCGCAT ACACCAGAAA ATTACGTGTG GCCGATTGCT TTAGCAATGG AAGGAATGAC	1080
AACCAAAGAT AAAGCAGaAA AAGAACGAAT TTTAGATTTA TTAGTTGCCA CAGATGCGGG	1140
GACTCATTTG ATGCATGAAG GTTTTGACGT AGATAATCCT GAAAATTATA CAAGAGAATG	1200
GTTTTCATGG GCTAATAaGA TGTTTTGTGA ATTAGTTATG GATTATTTTG ACATCCGAGT	1260
GGAAAAATAG GAGGAAAAGT AATGACAAAG AAAAAAGTCT ATATTGTTTC CCATAGTCAT	1320
TGGGATCGTG AATGGTATCT GCCTTATGAA GaACATCATA TGCGCTTAAT TGAATTAGTG	1380
GATAATGTAT TAGATTAAAT TGAAAATGAT CCAGAGTTTA ACAGTTTCCA TTTAGATGGA	1440
CAAACGATTA TCTTAGACGA CTATTTACAA GTTCGCCCAG AGAAAAAGA AGCCGTGAAA	1500
AAAGCTGTCC AAGCAGGCA ACTAAAAATC GGACCATTTA GTATTTTACA AGATGCTTTT	1560
nAATTAGCTC TGAATCCAAT GTCGAACATG CTGATTGGGC TTAGAAAGCa AAATnGGGGC	1620
ACAGTCAATT AGGTACTCCC G	1641

## (2) INFORMATION FOR SEQ ID NO: 121:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8033 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

AGAGAAGGGA TCATATGTCA GCAAAACATA ATCATGAACA ACACCAGGAA AAACAGCCAA	60
CAATGTCCCA CGATAAAATG AATCACATGA ATCATGATAT GGCGCATGAA CAAATGGCTA	120
TGTCACACCA ACATGAAGGA ATGAAGGGAA TGGATCATTC AATGCATATG GGGAATTTTA	180

AACAAAAATT	CTGGCTGTCA	TTGATTTTAG	CTATTCCGAT	TATTGTTCTT	TCACCAATGA	240
TGGGTTTTCA	ATTACCTTTT	CAATTTACTT	TTCCAGGTTC	TGATTGGTTG	GTCTTAATTT	300
TAGCAACTGT	GCTATTTTTT	TATGGTGGGC	AACCATTTTT	AAGTGGCGCT	AGAATGGAAT	360
TACAGCAAAA	GAGTCCAGCA	ATGATGACGT	TAATTGCGAT	GGGGATTTCG	GTATCATATT	420
TTTACAGCTT	ATATGCGTTT	TACATGAATC	ATTTTACGAA	TCAGGCCAC	GTGATGGACT	480
TTTTCTGGGA	ATTAGCCACT	CTGATTGTTA	TTATGCTTTT	AGGCCATTGG	ATTGAAATGA	540
ATGCTATTTT	TAATGCTGGT	GATGCCTTGA	AAAAAATGGC	TGAACTGTTG	CCTGACACAG	600
TCAAGCGAAT	GACTGAACAT	GGGGAAGAAG	AAATTCCTTT	GCAAGATGTC	CAAGAAGGTG	660
ATCGCTTGAT	TGTCGTTTCG	GGAGATAAAA	TTCCGACAGA	TGGTAAAATT	TTGAAAGGAA	720
GCACAACAGT	CGACGAATCG	ATGGTTACAG	GGGAATCAAA	AACAGTAGAA	AAAAACATTG	780
GTGACTCCGT	GATTGGTGGA	GCTGTCAATG	GCAACGGAAT	GATTGaAATT	TCGGtAACTG	840
GCACGGGTGA	AAATAGTTAC	CTGTCAAAAG	TTATGGAGAT	GGTTAAACmA	GcGCAATCGG	900
AAAAATCaAA	ACTAGaATCC	aTTTCAGACC	gTGTAGCCaA	GTGGTTATTC	TATATTGCCt	960
TATTCGTAGG	TGTGTTAGCA	TTTATCGGTT	GGTTAATGGC	GACGAAAGAT	TTGTCGCTAG	1020
CATTTGAACG	AATGGTTACA	GTTTTTATTA	TTGCCTGTCC	CCATGCGTTA	GGCTTAGCAA	1080
TTCCTTTAGT	GATTGCTAGA	AGTACCTCCA	TTGCTGCTAA	AAATGGCTTA	CTTTTAAAAA	1140
ATCGGAATGC	ATTAGAACAA	GCCAATAAGG	TTGAATATGT	GTTGCTGGAT	AAAACCTGGAA	1200
CTTTAACAGA	AGGACAATTT	ACAGTTACTG	GCTTAGAATT	AATGAGTAAG	CAGTTTACTA	1260
GAGAAGAAGC	ACTTAAATAT	ATTGGGGCAT	TGGAAAAAAA	TGCAAATCAT	CCGTTGGCAA	1320
TTGGGATCAT	GAATATTTA	AtAAACAAGC	CgTTCAACCT	TATGAAGCCC	ATAATCTGCA	1380
AGCTTTATCA	GGCGTAGGAC	TGGTTGCTAC	TGTTCAAAAC	CAAGAAGTCA	AAATAGTCAA	1440
TGAAAAAGAA	GTTGCACGGT	TACAGTTAAC	CTTCGATGAA	ACAATTAAAA	CAAACCTATCA	1500
AGAGCAAGGA	AATACCTTGA	GCTATTTAAT	CATCGCTGGT	CAATTAGTCG	CACTGCTTGC	1560
TTTAGGAGAT	AAAGTCAAAC	CAGAAGCTAA	AACATTTATT	GCAGAGTTAC	AAGCGCAAGG	1620
AATTACGCCA	GTCATGTAA	CAGGGGATAA	TCAAACAGCC	GCCAGTGCAG	TTGCTAATTA	1680
TTTAGGAATG	AAGGAATACT	ACGCAGAGCT	CTTACCAGAA	GATAAAGAAA	AAATTGTGCA	1740
ACAATATCTT	ACTGAGGGAC	ATCAGGTCAT	GATGGTAGGA	GACGGTATTA	ATGATGCACC	1800
AAGTTTAGCC	CGTGCTTCAA	TCGGCATTGC	GATAGGCGCT	GGTACAGATG	TAGCGATTGA	1860
TTCCGGCAGAT	GTTGTAAGTA	CGGATAGTGA	TCCTAAAGAT	ATTTTAAGGT	TCTTAGATTT	1920
AGCCAAACAA	ACACGACGTA	AAATGATCCA	AAATCTTTGG	TGGGGAGCTG	GGTACAATAT	1980
TGTCGCCATT	CCTTTAGCGG	CTGGGGTTCT	TGCACCAATT	GGGATTGTTT	TAAATCCAGC	2040
TGTAGGTGCT	GTCTTGATGT	CATTAAGTAC	AATTATTGTA	GCAGCAAATG	CTATGACACT	2100
GCaTATTTcA	AAAAAATAAT	TCGATaTAAT	TGTaGAGAGC	GGAAAGTAAT	CAATACTTTT	2160

TGTTCTCTAT TTTTTCTTC TTAAAAAAT TTTGCCTTA TTTGACGTAT TCTTATwACa	2220
AGGGAGTGaA AGGGGACGAT ATATGGaTTT CaAGGaATTA TTGgTCTTTk GAmCATAAAA	2280
ATATGTTGGA ATTGTTGGGA TTTACTGCCA ACAAGAGCAG ACATTTTCTG TTTTGGATGG	2340
CTGCCAGAAA ACTAAAGGTG AGCGACAAAA CCTCCATTGT TACATTCAAC GTTTTGAAAA	2400
AAAGCAGGAA ACGTTTAAGA TAGAAGGGCT TTGTAAGATT TATACAACGA CCAAAGGAAT	2460
TGTAAATATAT CGTGAATTAT GCGTTTGTGG CTTGGCTCGA TTTAGACAAA AATTTTGTTA	2520
TTTTATTCCC GAATTTTATA TTTTACGTTG TCTAATTGAA GAGCGAATCA ATTATGCACA	2580
ATTGACCCAG GTATTGGAAT TACAGGAATC AACTCTTCGG AAGAAATTTT CTAATATTCTG	2640
ACGTTGGCTA GTTAATTTTG ATATTATTGT ACGTCAAAAA CATTACGATT TAACTGGTAA	2700
TGAATGGCAA ATTCGACAGT TAATCCTTTG TTTTATTTG TTTTCCAAG AAAGCTGTCT	2760
GGAAGAAAAC AGAGAAATGA CTCGAAAAT CATTACCTTT TTTGAGTTGG ATTTAAATGT	2820
AGCACAACAG AATCATCTTA GTTGGCTAAT TTATATTTGG GAGAAACGAT ATAGAGGCGG	2880
GCATGAAATT TCAGTGCCAA ATGCTAATTT ATTTCAACAA ACAAGTGCTT TTTTTTATTT	2940
ATTTCTGTGT GAAGTGTTGA GCACTTCTTT TATGTCTCTT AAAGAACAAA AAGCGTTATT	3000
TGTGATTCTT GAAGCACATT TTGGTGGTTG TTTTGGCAAG AGAGCGCGAA AGTATTTTAT	3060
TCATGAGCAA ATGAAAATAG AAAGTTTATG CTAAAAACA GCTATTTTTA TTATGAAAGA	3120
AATTAGGAGA AATTTTACAC AACATCATTT TAATTATCAA GAAATTCATT TATGTCGATT	3180
TTTAAGTACG CATATGAATA GTTTATTAGA CGGACAAGCA TGGTTGCCAG CTCATAAACA	3240
GGAACAAACA CTCGCTGCTC GTTATCAACA AACCTGGCAC AGATTACAAA AACTTATTAG	3300
GCTTTTAAAA CGCCTATATC CAGTCTTTAC TTCAGTAAAA GAACGGAAT TGACGAGTTG	3360
CTATTTTTAT CACATCCTAG ATTTATTTAA CCCGATTTTA TATGAAAAA AATATATTAT	3420
TTGCCTATTG ACGGACTTTC CTCCAGAAAA GGAACAAGCG TTGGGGCAAT CTATAAAAAG	3480
CTATTTTCAGT GAAAAAAGA ATATTACAAT TATTCATGGA AAACCAACGT ATCAACTTCA	3540
CCAAGTACAC CTCTTAATTG TCAATCATTT GTTTCAGATG AACGTAGCGC TTTCTTCTAA	3600
GACAGTGGTT TATTTGCCAG AGGAATTATC TCCTGCTTTT TTTGAAAAAG TCGAAGCGAA	3660
TCTATCGTAA GAAGTCCAAT TTTGTCTAAA TAGTGTTAAA ATGAAGAAGT AAGAGGAATG	3720
GATGAGGAGG ATATTAAAAT GAAACAACTG cAATTTGGTA CAAGTGACGA AACAGTTTCT	3780
TCGGTTATCT TAGGATGTAT GCGCTTAAAT GGTGCTGAAA ATCCTCAACA GGTAAATTGAA	3840
ACTGCTTATG ATCATGGCAT TACATTTTTT GATCACGCCG ACATTTATGG TGGCGGCGAC	3900
TGCGAAACCA TTTTGGTAA GGCCTTAAAA GAAAGTACCA TTCGTCGTGA AGATATTTTT	3960
ATACAAACGA AATGTGGGAT TCGTCAGGGC TTTTGTGACT TTTCAAAAGC ACATATTTTA	4020
GAAGCGGTCG AAGGGAGTTT ACAGCGTTTA GGTGTAGATT CAGTCGATGC ATTATTGTTG	4080
CATCGTCCAG ATACTTTAGT GGAACCAGAA GAAGTAGCTG AAGCTTTTCA TTTATTAGAA	4140

AAACAAGGGA AAGTTCGCTA TTTTGGTGTA AGTAACCAAA CGCCAGGACA AATTGAACTT	4200
TTAAAAACAG CTGTAAACA ACCGTTATTA GCCAATCAAT TACAATTTGG CATCAAGCAT	4260
ACTGGTATGG TAGATCAAGG CTTACAAACA AATATGGAaA TATCAGGAAG CATTGATTAT	4320
GATCATGGTA TTTTAGATTA TTCACGGTTG AAACAAATGA CTATTCAAGC ATGGTCACCG	4380
TACCAATATG GTTATTTTGA AGGTGTCTTT ATTGGAAATG AAAAATTCCC TGAGTTAAAT	4440
CAAAAATTAA GCGAATTGGC TGAAAAATAT CAAACAACGC CAACTGGTTT AGCCAGCAGC	4500
TGGATTCTGC GCCACCCTGC TAATATGCAA GTCATCGCAG GAAGTATGAA TCTTGGTAGA	4560
ATTGAAGAAA TTGCTAAAGC AGCAGACATT GTGATTAGTC GCGAAGATTG GTACGATATT	4620
TATCGTGCAG CGGGAAATGT TTTGCCGTAA AAAGATTCTC ATACTTGTGA AGATCCAGCG	4680
TGTATTGCTG ATTGATTTTG GTATTTGATA TAGTTGTCAC TAAGGAACTA TGTCAAATAC	4740
CTTTTCTTT CATTCTGCAA TCGCCGTCTA CTGCAAATTC TGTAGTTTTT TGTTAAATA	4800
GAAAGGAAAC TAGTTTGCTA AAGATGAGGG GAAAAAATG ACCAAAAAA AATTATTACT	4860
CGTTGATGGA AACAGTGTAG CCTTTCGTGC ATTTTTTGCG CTACATAACT CATTAGAACG	4920
ATTCAAAAAT AAGAACGGCT TGCATACAAA TGCTATTTAT GCATTTAACA ATATGTTTGa	4980
AAATGTAATG cAAAAAGAAA TGCCTACTCA TGTTTTAGTC GCTTTTGATG nCAGGGAAAn	5040
CCACATTcAG AACaGrATTT TATTCaGAAT ATAAAGCCGG ACGTTCAAAA ACACCGGGGG	5100
AATTTAAAGA ACAAATGCCG TATATTCGTG AACTTTTAGA AGGTTTAGGC GTAAAAATATT	5160
ATGAActACC TAACTATgAA GCAGATGATA TTATTGGTAC ATTAGnCCAA CAAAGTTGAT	5220
AAAGATGAAT TTGATGTTGT CGTGTTATCA GGGGACCGAG ATTTGACGCA ATTAGCGACG	5280
GACACTGTCA AAGTAGATAT CACTGTTAAA GGAGTTAGCG ATATTGAATC ATATACGCCT	5340
GAGCATATTG CTGAAAAATA TGATGGGTTA ACGCCTAAGC AAATTATTGA TATGAAAGGG	5400
TTGGCTGGCG ACACGTCGGA TAACATTCCA GGTGTAACGA AAATCGGTGA GAAAACCGCG	5460
ATTAAGCTGT TAAAACAATA TGGCTCAGTT GAAGGGGTTT ATGAGAATAT CGATGAAATG	5520
AAAAAAGCA AAATGAAAGA AAActTGATT AATGATAAAG AGCAAGCCTT TCTTTCTAAA	5580
CGTTTAGCGA CAATTGAAGT CAATGCACCA GTCGAAGTCA ATGTGGaAGA CTTGGcATAT	5640
GAAGGGAAAA ACTTAGaAAA ATTGGTTCCCT TTTTATAAAG AAATGGACTT TAAACAATTT	5700
TTAGCTAAGT TGGACATTAC AGAAGAACCT GTTGAAATGG AAGATATTTT ATTTGAAGTA	5760
GTCGAAGATC AGTTAACCAA TGAAATGTTT ACAGACGATA TGGCTCTTTA TGTGAGATG	5820
ATGGAAGATA ACTATCATAC GTCACCGATT GTCGGCCTTG CTGGGGCAA CAACAAGAAG	5880
ATTTATACCA CGAATAATTT AGCTGTTTTT GAAAGTCAGC CATTTATTGA CTGGTTGATG	5940
GATGAGACAC GTAAAAAAA TGTCTACGAT GCaAAACGGA CGTaCGTTGC ACTAAATCGC	6000
TATGTAGGAA AAATGACAGG GATTGCcTTT GATGTTTTAT TGGCTGCTTA TTTACTTGAT	6060
ACCAATGACA ATAACGCTGA TATTGAAGGT GTAGCACAAC ATTACGGATA TGATGCGATT	6120

CAATCTGATG AAGCCATTTA TGGAAAAGGG GCCAAAAAAG GCTTGCCAGA GGACGAAGAA	6180
GTTTTCTTTG GTCATTTAGC ACGTAAGATT AAAGCAATTC AATTTTAAAC AAGCAAATTA	6240
GATAGCGAAT TAACAGAAAA AAATCAAGCT GATTTATTCT ATAAAATGGA ATTACCACTT	6300
TCTCGTATTT TAGGCGATAT GGAAATAACT GGTATCCGTG TTGATGCCAC ACGCTTAAAA	6360
gAAATGCAAG TAGAATTTTC AGAACGCTTG AAAGAAATCG AAGAAAAAAT CTATGCAGAA	6420
GCTGGCGAAG AATTTAATTT GAATTCACCT AAACAATTAG GCGTTATTTT ATTTGAAAAA	6480
ATGGGCTTAC CTGTGATCAA AAAGACGAAA ACTGGCTATT CAACAGCGGT AGATGTTTTA	6540
GAACAATAA AAGAACAAGC ACCAATTGTG GCCGATATTT TAACTTATCG TCAAATTGCC	6600
AAAATTCAAT CGACCTATGT TGAAGGACTT TTTAAAGTAA TTCAGCCAGA TAATAAAATT	6660
CATACTCGGT ATGTACAAAC ACTGACaCAA ACAGGACGAT TAAGCTCAGT AGATCCCAAT	6720
TTACAAAACA TTCCTATCCG TTAGAAGAA GGCCGAAAAA TTAGAGAAGC GTTCGTACCA	6780
AGAGAGGATA ATTGTTGAT TTTCTCTTCT GATTACTCTC AAATTGAGTT ACGTGTGCTC	6840
GCACATATTT CTAATGATGA GCATTTAAAA GAAGCATTTG TGGGAAGGACA AGATATCCAT	6900
GCCAGCACTG CAATGCGAGT TTTCGGTGTT GAAAAAGCAG AAGATGTCAC ACCTAACATG	6960
CGTCGTCAAG CTAAAGCGGT GAATTTTGGG ATTGTCTATG GTATCAGTGA TTATGGCTTG	7020
TCTCAAACT TAGGCATTAC TAGAAAGCAA GCGCAACAAT ATATTGACAC CTATTTTGAA	7080
AAATATCCAG GCGTGAACA GTACATGGAA GAGATTGTCC GCGATGCCAA AGACAAAGGC	7140
TATGTGGAAA CATTGTATCA TCGCCGTCGT TATTTAAGTG ATATTAATTC AAGAACTTT	7200
AATCTACGTT CGTTCGCTGA ACGTACAGCG ATTAATACAC CGATTGAGG AAGTGCCGCA	7260
GATATTTTAA AAATTGCGAT GATTGAAATG GCGCGTCGTT TAAAAGAAGA AAAATTACAA	7320
GCGACAATGT TATTACAAGT GCACGATGAA TTAGTTTTTG AAGTACCAGA ATCAGAATTA	7380
GAACAATTAA ATCAACTCGT CAAAGAAGTG ATGGAGCACG CAGTTTCCTT ACATGTGCCA	7440
CTTATCACCG ATAGCAGTTG GGGAAAAACA TGGTATGAAG CAAAATAATA GCTTACCCAT	7500
TAAGAACGAT GAATAAAATA AGGATAAGTA GGATCTGGAC GTTGATATCT AAGCGTTCAG	7560
ACCTCTTTAT ATAAAGGAGC AAAAATAATG CCAGAATTAC CAGAAGTAGA AACAGTGCGA	7620
AAAGGACTTG AAAAATTGT TGTAGGAAAA ACCATTCAAG AAGTCATTGT TTTTGGCCA	7680
AGAATTATTG AATCACCAGA AGTTGATGTA TTTCAAGTC AATTAGCTGG TCAAACGATT	7740
GAAGGGATTG AGCGAAGAGG AAAATTTCTG ATTTTAAAT TGTCTGATAA TGATATGATT	7800
TCTCACTTAC GGATGGAAGG AAAATATGAG TTTCAACAGG CTGATGATGA AATTGCCAAA	7860
CATACCCATG TAATGTTTAC TTTTACAGAT GGCACGCAAT TACGTTATTT AGATGTACGG	7920
AAATTTGGGC GTATGACATT GGTTCTAAG AATCAAGGTC ATCAATATAA AGGGATTTTA	7980
GCTTTGGGCC CAGAGCCAAC ACCAGACGTT TTCCAATTAG CTACTTTTCA ACA	8033

(2) INFORMATION FOR SEQ ID NO: 122:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9813 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

```

GTTTGGATAA TCGGCGTTAC TTTAGCAATC AGTGCTTTTC TTTAGCTAG CTGTGGGCAA      60
ACCACACAAT CGCCAAAAGA GAAGAAAGAG CTCACGGTAA TGACTACTTT TTATCCGATG      120
TACGACTTCA CAAAACAAGT CGTTGGTGAT GAGGGAGAGG TTGAGCTGTT GATTCCAGCT      180
GGAACGGaAC CACACGATTA TGAACCAAGC GCAAAAGACT TAGCGAAAAT TACTGATGCG      240
GATGTCTTTG TTTATaATAG TAAAGAGCTA GAAACTTGGG TTCCGAATGT TATTGAAAAC      300
TTGGACACGA AAAAAGTTTC AATCGTGGA GCCAGTCAAT CCATTCAATT GATGCAAGGA      360
ACAGAAGAGG AAGAATCTGG CGaAGAAGGT CATGAAGGCC ATAACCATT CCACGAACTG      420
GATCCGCATG TCTGGTTAGA TCCCGTGTTA GCGCAAAAAG AAGTAACAGC GATTCGGGAT      480
GCGCTTATCA AAAAATATCC TGAGAAAAAA GCCGTATTTG AAAAAAATAC CGTTGCTTAT      540
TTGGA AAAAAT TGACTGCGTT AGATAAGGAA TACCAAGCTG CGTTTGCAGG CGCTAAAAAT      600
CGGACCTTTG TGACGCAACA TGCTGCATTT GGCTATTTGG CGAAACAATA TGGCTTGACC      660
CAAGAACCGA TTGCTGGGAT TTCTCCAGAC CAAGAACCTT CGCCAAGTCG TTTAGCCGAA      720
TTAAAAAAT ATATTAAAAC AAACAATGTT TCTGTGATCT ATTTTGAAGC ATCGGCTTCT      780
ACAAAAGTAG CCAAAACATT GGCAGACGAG ACAGGCGTCG AATTAGCCGT ATTGAATCCT      840
TTAGAGAGTT TAACTCAAAA AGAACAAGAA GCAGGCGAGA ACTACGTGTC TGTCATGAAA      900
GAAAACCTCG CTGCTTTACA AAAAAGCATT CATTAACTCG GCAACAGAAA GGAGCCAACT      960
TATGCATTAT ATTGAAGTCG AAAATTTAAC CTTCTATTAT GATGATGAAC CTGTTTTAGA     1020
AGATGTTTCT TATTATGTTG ATCCAGGGGA ATTCGTTATT TTAACAGGAG AAAATGGCGC     1080
TGCCAAGTCA ACGCTGATCA AAAGTACCTT AGGGTTATTA AAACCAACAA GTGGCAAAAT     1140
TACGGTCGCT AAAAAAATA GTGCTGGTGA AAAAATCAGT ATTGGCTACA TTCCACAACA     1200
AGTCGCTTCT TTCAATGCGG GCTTTCCTAG CACCGTCATT GAATTAGTAC GTTCAGGTCG     1260
CTTTCCAGA AATCGTTGGT TCAAACCGTT GACGAAAAAA GATCATCTTC ATGTGGAAAA     1320
AGCTTTGAAA TCACTGGACA TGTGGGAAAT GCGCCACAAA CGAATTGGCG AGCTTTCTGG     1380
AGGGCAAAAG CAACGAATTA GTTTAGCACG AGTGTTTGCG ACCGATCCAG ATTTATTTAT     1440
TTTAGATGAG CCAACAACAG GTATGGATGA ACAATCGCGA AACGAATTTT ATCAATTGTT     1500
GCAACACAGT GCGCATGAAC ATGGAAAAGC TATTTTGATG ATTACGCATG ATCATGAAGA     1560
TATCAAAACC TATGTGGATC GTCAAATTCG TCTTGTCGCG AAAGAAGATT CGAAATGGCG     1620
TTGTTTCCAT ATGAGTGAAG AATCGTATAC GTAAATTTTC CAATAAAACG TTATCTGGCT     1680

```

CGATCATAAC	GAAAGCCTAG	GAACGGCTTA	AAGAAATGTG	GGAAAAATTT	AAAAAATAAG	1740
CAGGAATTAA	AAAAATAATT	CGCTTAAACG	TCGCTACAAT	GGTAGTGTTT	ATTTTAGCGG	1800
TAAGGAGTGA	ATCGTATGGC	AGAAATGCTT	TCTTATGCAT	TTATGCAAAA	GGCCTTTTTA	1860
GCAGCACTGT	TTATCTCAGT	GATTGCCCCA	ATGCTCGGCG	TCTTCTAGT	TATTCGCCGA	1920
CAATCTTTAA	TGGCAGATAC	CCTTTCACAT	GTGTCATTAG	CCGGTGTGGC	ACTAGGCTTC	1980
TTTTTTAATT	GGAATCCTAA	TTAATGACC	TTAATTGTCT	TGATTGTGGC	TGCAATCATT	2040
CTAGAATATT	TACGAATCAT	TTATAGCACC	TATTCAGAAA	TTTCGATTGC	TATTTTAATG	2100
TCAGGCGGTT	TGGCTTTGGC	GTTAGTTTGT	ATGAATTTAA	CAGGAGGCAA	TTCAGCTGCT	2160
AGTATTCAAT	CGTATTATT	TGGTCCATC	GTCACGATTA	CGTGGGATCA	AGTGGTTATG	2220
TTGGCAATTT	TATTCGTAGT	TTAGTTCTA	TTGTTTATGT	TATTTAAACG	TCCAATGTAT	2280
GTTTTAACAT	TTGATGAAGA	TACTGCTCAT	GTTGATGGGC	TACCTATTCA	TTGGATGTCT	2340
ATGCTTTTTA	ATGTAATTAC	TGGTGTGGCG	ATTGCTGTGA	TGATTCCGAT	CGCGGGAGCC	2400
TTGTAAATTT	CAGCAATTAT	GGTCTTACCA	GCTGCAATAG	GTATGCGAAT	TGGTAAAGGC	2460
TTTAACACGG	TGATTATTAT	CAGTGTGTTT	ATGGGCTTGA	TTGGCATGCT	AACAGGGTTG	2520
ACTAGCTCGT	ATTATTTGGA	AACACCACCG	AGTGCAAGTA	TTACCCTAAT	TTTTATTGGT	2580
TTATTCTTAT	TAGTCAATAT	TTATCGCCGA	GTGGTTGTCA	TGGTCCAACG	AAAACAAAAA	2640
ATGCAAAGAA	ACTAAGGAGA	AAAACGAAAG	TTTTTCTCCT	TTATTTCTTT	TTTTTAAAAG	2700
AAATGGCGAA	CTATTTTCAG	TTATTTGAAA	AAAATCCTCG	AATTTTTCGG	TTAAACCCCT	2760
TATAATAAGG	AACGATAGAA	TGAAAAGGAG	TGACGAATTT	GAAAATTCGG	AGAAGTGAAC	2820
GATTAATCGA	TATGACCCAA	TATCTTTTAG	ATCATCCCCA	TACATTGGTT	TCTTTGACAT	2880
ATTTTGCTGA	CCGATACCAA	TCTGCCAAGT	CTTCCATTAG	TGAAGATTTA	GCAATTGTTA	2940
AAAAAACATT	TAAAGAACGT	GGAACAGGTA	TTTGTAGAAAC	GATTCCAGGC	GCTGCAGgCG	3000
GTgTCCgCTT	TATCCCTGAA	ATTCCTTATG	AGGAAGCGGA	CAGTTGATTA	TGGATTTATG	3060
TGATCGTTTG	TCAGAACAAAG	ATCGCTTATT	ACCTGGGGGC	TATGTTTATC	TTTCAGATTT	3120
ACTTGGGGAA	CCGAATTTAT	TAAGACAAGT	GGGCCGGATT	ATTGCTTCTA	AATATTTAGG	3180
CAACAAATTT	GATGCGGTAA	TGACCGTTGC	GAATAAAGGA	GTGCCAATTG	CGCAAgtCTGT	3240
GTCTTATTAC	TTAAATGTTC	CGTTTGTGAT	TGTTTCGTCGC	GATTCTAAGA	TTACTGAAGG	3300
TTCAACGGTG	AGCGTGAAC	ATGTATCAGG	GTCTTCAGAA	CGTATCGAAA	AAATGGAAC	3360
TTCAAAACGC	AGCTTAAAAC	GCGGCTCAAA	AGTTTTAGTC	GTGGATGACT	TCATGAAAGG	3420
CGGCGGCACC	GTCAATGGTA	TGAAGAGTAT	GATTGAAGAA	TTTGAAGCAG	AACTTGTTGGG	3480
CATTACCGTT	TTTGCTGAGT	CAAAATTCAA	TGGCCGTCGT	GCAATCGATG	ACTATACTTC	3540
GCTATTGTAC	GTGGAAGATG	TGGATACACA	AACAAAAACG	ATTAGCGTAG	TCCCAGGAAA	3600
CTATTTTACA	GAGAAATAAT	TTGAAAAAGA	ACCTGAAAAT	GGTCCATTCC	ATTTTCAGGT	3660

TCTTTTTTTC TGTTTTATTC CTTTAAAAA AGATGAGAAA ATAAATTGGT CTACCTTTTA	3720
AATGCCTTAT TTTGTTGAAA AAGACAAAAG AGTAAGATAA ACTAGAAAAG GTATAATGAG	3780
GACGATTTAG AAGAAAATCA AACAAATTTAG TTAATTTGAG GTGGCAAACG TGGAAAATAG	3840
ATACGCAATT ATTTTAGCAG CTGGTAAAGG AACACGCATG AAATCTAAAC TTTATAAAGT	3900
ATTGCATCCA GTTGCTGGTA AACCAATGGT CGAACATATT TTAGATCAAG TAGAACAAAC	3960
AGAACCAACA GAAATCGTGA CAATCGTTGG ACATGGGGCG GAAATGATTA AAAGCCATTT	4020
AGGCGAACGT AGTCAATATG CCTTACAAGC TGAACAATTG GGAAGTGGGC ATGCAGTCAT	4080
GCAAGCACAA GAGTTATTAG GTGGTAAACA AGGAACAACA TTAGTTATTA CAGGGGATAC	4140
GCCGTTATTA ACTGCGGAAA CCTTGAAAAA TTTATTCGAT TACCATCAAG GTAAGAATGC	4200
AAGTGC GACT ATTTTAACAG CGCATGCGG <sub>a</sub> AGACCCAACA GGCTATGGTC GAATTATTCG	4260
TGATCATGTG GGCATTGTTG AACGAATCGT GGAACAAAAA GATGCCAGTG AAGAAGAAGC	4320
ACGTGTTCAA GAAATTAATA CAGGAACCTT CTGTTTTGAT AATGAATCAT TGTTTGAAGC	4380
GTTAgcGnAA ACAGATACAA ACAATACACA AGGGGAATAC TATTTAACAG ATATCATTGA	4440
AATTTTGAAA AAAGAAGGCA AAGCTGTCGC TGCTTACCAA ATGGCAGACT TTGACGAAGC	4500
AATGGGTGTA AATGATCGTG TTGCTTTATC CACTGCTAAT AAAATTATGC ATCGTCGTTT	4560
AAATGAAATG CATATGCGAA ATGGTGTTAC ATTTATTGAT CCAGACACAA CGTATATTGA	4620
TGAAGGCGTG GTAATTGGTT CAGACACAGT CATTGAAGCG GGAGTCACTA TCAAAGGAAA	4680
AACAGTGATT GGCGAAGATT GCTTGATTGG CGCACATTCA GAAATCGTTG ATAGTCACAT	4740
CGGCAATCAA GTGGTTGTTA AACAGTCTGT GATTGAAGAA AGTGTGGTTC ACGAGGGGGC	4800
CGATGTGGGT CCGTATGCAC ATTTACGTCC TAAAGCAGAT GTGGGGGCAA ACGTACACAT	4860
TGGTAACTTC GTGGAAGTAA AAAATGCAAC AATCGATGAA GGCACAAAAG TGGGCCATTT	4920
AACATACGTT GGTGATGCAA CATTAGGCAA AGATATTAAT GTCGGTTGCG GCGTTGTTTT	4980
TGTTAATTAT GATGGCAAAA ATAAACACCA AACAAATCGTG GGTGATCACG CTTTTATTGG	5040
CTCTGCAACG AACATTGTTG CGCCAGTCAC GATTGGTGAT CATGCGGTGA CTGCTGCTGG	5100
TTCAACCATC ACAGAAGATG TCCCTTCAGA AGATTGGCG ATTGCCCGGG CACGTCAAGT	5160
GAATAAAGAA GGCTATGCTA AAAAGTTACC TTATATGAAG GATTAATCTC AAATAAAAAA	5220
GAACGCTCTT TTGCTTAAGC AAAAGAGCGT TCTTTTTCCT CCTTAAAGAC TAGACCAATT	5280
GCTGCTAGTA TTTGTTTAGA GACGTTCAAG GACTAGACCA ATGCTGTTTA TTCGGTATTT	5340
TTAGTTATTT TTTTCAAAAA AATAGTTTTT GTTCCGCTTT TCCAATATAA AAAGACTTTC	5400
TTTTTATGTA AAGTTCGTTT TTTGGTCAAA ATTACTATTT TTTTCACAAA CAAATCCGCT	5460
ATATTGTCCA AAAATAAAAA CAATTTACAG GAAAAACATC TTATTTTAAC GGAATTTACT	5520
GTTTTTTTATT TAACGGATAT ATTATGATAG GAAAAGAAGA GCAGCGTGAT GGGGGAAGGC	5580
TCTTTTTGTT GTTTGTTACT TTGACAATTG TTTGTTGCAA ATGGGTTGAA ACACAAGGAG	5640



GAAATTTGTG AAAAAGTTGA GCTTTAAAAA AGTGAAGTGG GGCATGCATT TTTTAATGGC	5700
TGTTGCGTTG ATAGCGCCAA GTGTTACTAG TACGGCATAT GCAGTAGAAA CAACGAGTCA	5760
ACAAAGTTCA GAAGCAGTAA CAAGTACCAC CGATTCAAGT AGAAAACAAG AACCAGTCAT	5820
TACACAGGAA ACAACAGACA TCAAACAAGA AGCACCAAAT CAGGCTACGA GTGACAGTGT	5880
CAAGCAGTCA CAAGAAACCA CAGCACCAAC AGAGACGACG AATTTAGAAA CGTCAATCGC	5940
TGAAAAAGAA GAAACGAGCA CGCCGCAAAA AATAACAATT TTAGGTACGT CAGATGTTCA	6000
TGGTCAATTA TGGAATTGGT CTTATGAAGA TGATAAAGAA CTACCAGTTG GTTTGTCCCA	6060
AGTAAGTACA GTCGTTAACC AAGTCCGGGC AAAAAACCCA GCAGGCACCG TTTTAATTGA	6120
TAATGGCGAC AATATTTCAAG GCACTATTTT AACAGATGAC TTGTATAATA AAGCGCCTTT	6180
AGTGAATGAA AAGACCCATC CAATGATCAC CGCCATGAAT GTGATGAAGT ATGATGCAAT	6240
GGTTTTGGGA AATCATGAGT TTAATTTTGG TTTACCGTTA ATCAAAAAAA TTCAACAAGA	6300
AGCCACTTTT CCAATCTTGT CTGCGAATAC CTACAATAAG GAAGATGGTC TTCGTTTTGT	6360
TGAAGGGACT ACCACGAAGG AACTTGATTT TAATCAAGAT GGGCAGCCAG ATTTAAAGT	6420
TGGGATTATC GGCTTAACAA TTCCGCACAT TCCTTTGTGG GATGGCCCTC GTGTTACTTC	6480
GCTTAATTTT TTACCTTTGA AAGAAGAAGC AGAAAAAGCA GTTACTGAGT TGAAAGCTAA	6540
CGATCAGGCT GACATTATTG TTGCCTCGAT TCATGCGGGA CAACAAAATA GTGATCCGGC	6600
TGCCAGTGCC GACCAAGTAA TTGAAAATGT CGCGGGGATT GATGCGTATA TTCTGGGTCA	6660
TGACCACCTT TCTTTTACCA AGCAAGGAGC AGCGCCGAAT GGAAAACTG TACCGGTAGG	6720
GGGACCGAAA GATACGGGGA CAGAAGTTGT CAAAATTGAT CTTTCAGTTG CTAAAAATGC	6780
CGATAAGTGG GAAGTGCAAG AAGGTACAGC AACGATTGTA CCAACAACGA ATGTTCCAGC	6840
AGATGAAGCA GTTAAGGCAG CGACAAAAGA ATACCaTGAA AAAACGCGAG CGTTTATTCA	6900
GGAGGAGATC GGCACAGCAA CAGCTGATTT TTTACCAAAA CAAGAAATTA AAGGAATTCC	6960
CGAAGCACAA TTACAACCAA CAGCGATGAT TTCTTTAATT AATAACGTTC AAAAAGAAGT	7020
AACGGGCGCA CAATTAAGTG CGGCAGCGCT GTTTAAATAC GACAGTAAAT TACCTGCGGG	7080
GAAGATTTC TATGCCACGA TTTTGTATAT CTACAAATAC CCGAATACCT TAGTGAGTGT	7140
TCCCATTAAAC GGTGAAAAC TACTGAAGTA TTTAGAAAAA CAAGGGGCGT ACTATAACCA	7200
AACACAGCCA GATGATTTGA CCATTAGTTT TAATCCAAAC ATTCGTGTAT ATAACATGA	7260
CATGATTTCT GGAGTGGACT ACAAGATTGA CATTTCAAAA CCAGTGGGTG AACGAATTGT	7320
AGATGCGAAA ATTGACGGCC AACCGCTGGA TCCTGCCAAA GAATATACGA TTGCTATGAA	7380
TAATTATCGT TACGGCGGTT TAGCTAGCCA AGGGATTCAA GTAGGGGAAC CTATTAAAAA	7440
TTCTGATCCA GAAACCTTAC GAGGAATGAT TGTTGATTAT ATTAAGAAAA AAGGAACTCT	7500
TGATCCAGAA CAAGAAATCG AACGAAATTG GTCAATTATT GGGACAAATT TTGATGAAAA	7560
ATGGCGTGCC AAAGCAATCG AATTAGTGAA TGACGGCACT CTTCAAATTC CGACTTCTCC	7620

TGATGGACGT ACACCAAACG CCGCCGCTAT TACGAAACAA GATGTCCGTA ATGCGGGCTT	7680
TGATTTAGAT AATGCATATA CCATTATGCA CACAAATGAC GTTCATGGCC GACTAGAAGC	7740
AGGGAAAGgC GAATTAGGTA TGGCGCGTCT AAAAACCTTT AAAGACCAAG AAAACCCAAC	7800
CTTGATGGTG GATGCAGGGG ATGTTTTCCTA AGGATTACCA ATCTCCAATT TCTCCAAAGG	7860
CGCGGATATG GCCAAAGCAA TGAATGAAGT TGGTTATGAT GCCATGGCGG TGGGAAATCA	7920
CGAGTTTGAT TTTGGTTTAG AGATTGCACT AGGTTATAAA GACCAACTGA ATTTTCCGAT	7980
TTTATCTAGT AATACGTATT ACAAAGATGG CAGTGGACGG GTTTTTGATC CGTATACAAT	8040
CGTAGAAAAA TCCGGGAAAA AGTTTGCCAT TGTAGGTGTG ACGACCCAG AACAGCAAC	8100
GAAAAACAC CCGAAAAACG TAGAGAAGGT GACATTTAAA GACCCGATTC CAGAAGTAGA	8160
AGCAGTGATT AAGGAAATTA AAGAGAAGTA CGCGGATATk CAAGCTTTCG TGGTTACTGG	8220
GCATTTAGGC GTAGATGAAA CGACGCCGCA TATCTGGCGT GGTGATACGC TAGCAGAAAC	8280
CCTTAGTCAA ACATATCCTG AGTTAGATAT CACTGTGATT GATGGACATT CGCATACAGC	8340
CGTCGAAAGT GGCAAACGTT ATGGCAAAGT GATCTATGCT CAAACAGGTA ATTATTTAAA	8400
TAATGTTGGG ATCGTCACAG CACCAGAGAG TGAACCAACT AAGAAAACAA CAAAATTGAt	8460
TTCagcAGCA GAGCTGCTAG AATTGCCAGA AAACCCGGCA GTTAAAGCCA TCGTTGATGA	8520
AGCACGTACG AATTTTAACG CTGAAAATGA AAAAGTAATT GTCGATTATA TTCCATTAC	8580
ATTGGATGGA CAACGAGAAA ATGTGCGCAC ACGAGAGACC AACTTAGGGA ATTTGATTGG	8640
TGATGCGATT ATGTCATATG GCCAAGACGC GTTAGCCAA CCTGCTGATT TTGCAGTAAC	8700
TAATGGTGGC GGCATTCGCG CTGATATTAA ACAAGGGCCA ATTAAAGTTG GGGATGTCAT	8760
TGCTGTGTTA CCTTTTGSCA ATAGCATTGC GCAAATTCAA GTAACCGGCG CCCAAGTTAA	8820
AGAAATGTTT GAAATGTCTG TTCGTTGAT TCCACAAAAA GATGAGAATG GCACAAATTT	8880
ACTAGATGAT GCTGGCCAAC CAAAACCTGG CGCAAATGGT GGTTTCCTAC ATGTTTCAAG	8940
CTCCATTCTG ATCCACTATG ATTCCACAAA ACCAGGTACT CGCTTGGCTA GTGACGAAGG	9000
CAATGAAACA GGACAAACGA TTGTCGGTAG TCGCGTATTA GGAATAGAAA TTAAAAATCG	9060
GCAAACACAA AAGTTTGAAC CATTGGATGA GAAGAAACAA TACCGGATGG CTACCAATGA	9120
TTTCTTAGCT GCTGGTGGTG ATGGTTACGA TATGCTAGGT GGTGAACGAG AAGAAGGGAT	9180
TTCCTAGAT TCTGTCTTAA TTGAATACTT GAAAAGTGCA ACCAGCTTGC GGTGTATCG	9240
TGCAGCAACG ACGATTGATT TAGCACAATA TAAAGAACCA TTCCAGGCG AACGAATTGT	9300
TTCTATTTTC GAAGAAGCTT ACAAAGAGTT AATCGGTGGA GGAGAGACGC CAAAACGAGA	9360
TCCAAAACCA GACCCGAAAC CAACACCAGA AACACCAGTA GCAACCAATA AACAAAACCA	9420
AGCGGGAGCA AGACAGAGCA ATCCATCCGT AACAGAGAAG AAAAAGTATG GCGGCTTTTT	9480
ACCTAAAACG GGTACAGAAA CAGAAACGCT TGCATTATAT GGTTTACTGT TCGTTGGACT	9540
TTCTTCTTCT GGCTGGTATA TTTATAAACG ACGTAACAAA GCTAGTTAGT CATTTTTTAA	9600

739

GAACCGTCcT TCGGGATTGG TTCTTTTTTTT ATTTTAATGC TTGCCATTG TTTTAAATTT	9660
TAGTATTATA AGAATATtCT AAAAFTFAAA AAATGAAGAG TTTTCTGT GAAGGGAGAA	9720
GGGCTGAAAG TAAACGTATA AAGGACTTAT CGGGTTAAGT TTACAACGCT GTTAGTTTAG	9780
CTGCCTGTGG TGGTGATGGG GGACCTGGCG CCG	9813

(2) INFORMATION FOR SEQ ID NO: 123:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32768 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

CATTGGGAAA AGTGGAAGC CTAAACCTC GCAATCGCCC AAGGTTTACC AGCTGAACTC	60
GGCATCACTA AGCCgAAAAT TGAAAGCTGG AcCAACGGCT GGAATTTACG CAgTCATTTC	120
TGGGCTGCGT ATCGCAGTGA GCAACGTCAA GCGGAAAATG CTTGTTTGGC GTTGTTATTA	180
AATAAAAAGC AATTTCAGT GTATTTGATG TTTCAACATT ACAAAGTGA GGAGCGGACA	240
GGCTCAGTCG TCGCTTACAA TCAATTGCTT AATCGTTTAG AAGCTTGGAG TCAAACCATC	300
GACTGCGAAG GCTATTATAT TTGGCCGAG gAGGAGCATG AACTAGTGA CCATTTGCCT	360
TTAAAAGATT ATTTAGCAGA TGAAAAACAA CGGAAAATTC TAGAAGAGCG GATAGGCAAT	420
CGCTCATTTT AGATCGGTAA GTTGTGGTTT GCGACGGAAA AGCTTGAAAC GATTGAAGAA	480
AAAAGTTGTC AGGCGCTTCA AGAATTAGCG CCGTTGTATA ACGCATTATC TGAAAAATGA	540
ACGAAACAAC GTCTCTGACC AATTACCTGG TCAGAGACGT TGTTTGTGTT TTTATTGATT	600
TAAATCTGTT AGTAGTAGCT CGTTAGAAAC AGCAAAAGTA GTAACCTCTG TTAAAGCTGG	660
ATCCAAAGTG TATTTTGTA ATGAGGTAGT ACTTCTGGT TGGACATACA TGGTTAAATT	720
AGTTACATCA TATCCTGACC AATAATGTGT GAAATCAGAC AGAGAATTAC TTTGACCTAA	780
ATGGCCAATT CCCCAGAGAA TCATCACTGT TTGGAAGGCG CGATAGAGTT GCATAATCCC	840
TTCAGAGCTA TCAAATCTT GGAGATTGTC AGAATAGTAT AACGCACGAA TGAAGCGTGA	900
CGGCGAAGTG AAGTCTCCTG GGATTCTTAA TAGTCCGTAA GAAGTCCCAC CTTCAATTGG	960
GTCAAGTGT GCGCCTGATT TTAAAGTGTA AGGTTCCTTA ATAGCATAGT TGCGCATCCC	1020
gATATAGTTT CGTGCAATTG TTAAATGATA ATTGTATTCT GGACTGTTTG TGA CTACGCC	1080
AATCGAATCA TAAATGACGA ACGCTCCGTT ATCTGTTGGT TGCAAAACGA TACTATTATT	1140
ATTTTATCA TTAATGGTCC AGTGTAAGGT AAAATGAAGT TCCCCAATG AATCAAGTTG	1200
GAATTTTGA TCTAACAGGC CGATTTTTTC AAAAGCTACT TTTACTTCGT CAACGCTTTC	1260
AAAATTACTC AAAATATAAG CAAGTACTTC TTCTGCAATA TGAGGTGTTA AACCACGATC	1320
TGCTAAACTT TCAGCAGGTG CTGTTGAAGA TTCCATTAAA TATTGAATAT CACCAGATAA	1380
GCCGGCTTCA TTGATACCAT CTGTGATAGC AAGGCTGACG GTTTTACCAT CGTTCGCTGG	1440

TTGATCTGTG CCTGACATTG CTAGTCCCAT AAAGGCATAT TTTGCCGTCC AAGGATTTAG	1500
TTGGCTGTTT AGAACTGTAC CTTTTGGAAA TTGAGCAATT AGAGTAGGGA TTTTGGTG	1560
AATTGGATCT TCGTTGCCAA AAAAGTCAAA CGTAAAATCC ATCGTACGTC CATAAAAAAT	1620
ATCATTGTGTT TTGGAAATAA TCTTGATGCC TGTACACATA ATAGTTTCCC CAAATTAATG	1680
TAAGTTGTG AAGCCTGTTT ACTAGTTAAG CATATCATAC TCTAGGAAAC AGCGGTGGCT	1740
AAATTCCTAA AATATTAGAC AATTTGTCT AAATGAGGTG AAAAAAATTT TTTAGCGATT	1800
TCCTTCACAA AACACCTTAT TATCTTCCG AAAATATGAT ATCTTTAGAA TGGAAATTC	1860
ATGAAAAGAG GATGCGAATG TTTAAAAAGT ACACACAAAT GATAAGGGAT GAATTTAAAG	1920
GGTATAATCA AAAAAAATG AAAAAAGATT TCTTgGCaGt TTGACTkTCG CAGCCGTCGC	1980
GTTGCCATTA GCGTTAGCGT TTGGGGTCTC TAGCGGTGCT ACTGCAGCAG CTGGCATGAT	2040
CACGGCGATT ATCGCTGGAA TTGTGATTGG GAGCCTTTCG GGTGGGTTTT ATCAAATTC	2100
AGGACCAACA GGTGCGATGG CAGCTATTTT AATGTCAATT GCGGCAGTGC ATGGCATGCA	2160
AGGGATTTTG TTGGCCACCT TTTTGGCGGG GATTTTCTTG CTTTTGGCAG GGATTTTGC	2220
CTTAGGGAGC TTGACTTCTT TTATTCCGTC ACCGGTCATT ACCGGGTTTA CTCAGGAAT	2280
TGCTATTATT ATTGCGTTGG GGCAGCTTGA TAATTTCTTT GGGGTTcTT CAGAAGGACA	2340
CAATTTAATT GAGAAAATCA GCAGTTATCG TTCATTGGGC TTTGATTTAA GTATACCAAC	2400
GACGGTTATG TCGATTTTAG TGGTGCTTGG ATTAGTGTTT TTTCCGAAAA AGTGAATGC	2460
GATGATACCA GCTTCATTGG TGATGATTAT TTTAGCTACT TTAGCAACAG TGGTTTTCA	2520
ACTGCCAGTG GCCACTGTAG GGAAAATTCC GACAAGTATT TTTAGTGATA CACGTTTGA	2580
ACTTGGTTCA CTTCAATTAA GTGCGTTTCA GGATATTTTG GTGCCTTCGA TTAGCATTGC	2640
TATGTTAGGT ATGATTGAAA GTTTGCTTTG TGGCGCTTCG GCTGGTCGCA TGACAGGGAA	2700
ACCATTAGAC AGTAATCAGG AACTAGTGGC ACAAGGAATT GGCAACTTAA TTTTGCCATT	2760
TTTCGGTGGC GTGCCTGCGA CGGCAGCGAT TGCCCGGACG AGCGTAGTGA TTAAATCTGG	2820
TGCGCAAACA CGTTTAGCGG GTATTTTTCA TGCAGTCTTT TTATTGTTAT CGATGTTGCT	2880
TTTATCGCCA GTCATGTCTA AAATTCCCAT GCCTGCGTTA CTGGTGTTTT aGTGGTGACA	2940
GCGTGGCGAA TGAATGAATG GTCGGTGATT AAAGAATTCT TCCAGAAACG TTTCCACACA	3000
GCTTTATGTT TATtCTTCTT AACGATGGTA GCGACCGTTA TTTTGGACTT GAGTTTAGCG	3060
ATTGTtTTAG GGGTCTTAGC GGGTTGCCTG TTTTyCATTG CGAAAAGTGC CGTGATTACG	3120
ATTAATGTCG AGGAAGTTGA TTGGTCAAGA ATGGGGCTAG TACCGACCGA TTGTACGGAA	3180
AATTGGGTGG TAGTTTATTT GAGTGGCCCG TTATTTTCA TGTCTTCTGA ACGATTGAAA	3240
CAAACGTTAC TGCAATTGGA AACAAAAAGT GGTGTAGTAT TTTCAATGCG AGGAGTTCCT	3300
AGCATTGATA GTACGGCGTT AGATATTTTA GAAGAATTTT GTCTCGCAGC AGAAGGTCGT	3360
GGACAACAAG TTCAATTTGC CGCTCTGCAA CCAGAAGTTG AAAAAATGAT TCGAACCATT	3420

CAAGGAAATC AAGAGAAAGA ATACCATTTC TCAGTCGCAG AATTTTTACA AACGGTGCAT	3480
GTTGCAGAAC AATAATTCAT TTTCAAATGA AATGTTCTAT CTTCTATCGC GGGCTTAATG	3540
GGTGAAGAAA TAAGAAAAAT AAGTGGGGAC TAAAATGAGT TTCTAGAGGA GTTCCTAAAA	3600
AATAATAAAG AGAAACTTC TCAGTATGAA GTTTTCTCTT TATTATTTTT TTGGTTATTT	3660
ATTGAATTAC TTCTAGCATA ACGTCTTTTT GATCATCCGT AACATTGCCA GCGTAGTTAG	3720
GAGTGATTAA AGATATCGAA CCATTTTTAT TGTTGAATAA GTATAAAGTG CCCGATCTAT	3780
TTTGTTTCATC GTTTGAAGAA ATATTAGTTC CTAATAATTAA TTCTGTATTT ACTTTAACGG	3840
TGCGAATACT GTTATCGGCC GCTGAAAATA TTCTAATCTC TTTGTAGGA ATTGTATTTA	3900
TAGATAGCTC ATATTTATTA ATGGCTTCTT TCACTTGACC ATTTTCAGCC CCAAACAATT	3960
TAGTTCGGAA GGTAGCAGTT GGTGTTGAAT TTAAGTTCTC TAACGTAATT GAAGTTGGCA	4020
CATTCATTCC TTTAAATTA AATACGAGTG GATTGTTTAA GCTACTAAGA TCGACAGCAT	4080
AAGGTGTTGC TTCTAAAGAA TTTGCTTCTT CGACATTTGT AGATACTTCA TTTTTCGACT	4140
TTGACGAAGT TTGAGTGGTG CTGGCACTTT CTGAAGTTTC TTTTAGTTTA TTATTAGAGC	4200
TTGTTTCTAT GCTAGAACTG TAATGATTAG TACTGCTTTT TATTGATGTT GTATGATTGG	4260
TAGAATTTTC TTGTTTTTTA GGTTCATTCT CTTTTGAGA ACTACAACCT GTTAAAAAAA	4320
TAATAAAAT AACACAACT AATTTTTTCA TACTATTCCT CCTATAGACG CTATAACTAA	4380
TTTTATCTTA CTTGTTAATA TAACAAAAGA AATAATTATT TTATTTTCAC GAATGCTCAA	4440
ATTGTTTAA TCTGTAGTTT TTATACTTAT TTTTGTGCA TTTGCAACAA AAATAAGTTA	4500
AAGTAATGAC AATCATAAAA TATTAGCGAA AGGAAGTGTC CTAGTGACAG ATATTTTAAG	4560
AGAAATTGGC ATGATTGCGC GGGCCTTAGA TTCAATCAGT AATATTGAAT TTAAGGAATT	4620
ATCTTTAACA CGAGGACAAT ATCTTTATTT GGTGAGGTC TGTGAAAATC CTGGGATTAT	4680
CCAAGAAAAG ATCGCCGAAC TGATTAAAGT TGATCGAACC ACAGCGGCGC GGGCGATTAA	4740
GCGTTTGGA GAACAAGGCT TTATTTATCG TCAAGAAGAT GCAAGTAACA AAAAAATTAA	4800
GCGGATTAT GCCACAGAAA AAGGGAAAAA CGTTTATCCA ATCATTGTGC GCGAAAATCA	4860
ACACTCGAAT CAAGTTGCTT TGCAAGGCTT GTCAGAAGTG GAAATTCAC AATTGGCAGA	4920
TACTTAGTC CGGATGCGTA AAAATGTCAG CGAAGATTGG GAATTTGTCA AAAAAGGGAA	4980
CACTCGAAAT TATTAAGGAA AGGACGACTG ATTGATGGGA GAAATAAAC AATTAACAAT	5040
AGAAGATGCA CCAGCGTTAC AAGAAATCAG CATTGAACT TATACAGATA CTTTCGGACC	5100
ATATAATACC CCAGAAAATA TGCAAGCCTA TTTAGTAGAA GCCTACAATT TAGAAAAATT	5160
ACAAAAGGAA TTGGCTAATC CCCACTCTGA CTTTTATTTT ATCTTTGTCTG ATAATCAATT	5220
AGCTGGCTAC ATGAAATTAA ATCGTGAGAC GGCACAAACG GAACCGATGG GGCCAGAAAA	5280
GCTCGAGGTT GAGCGATTGT ATATTTTGCC TGCGTTTAAA GGaAAGAAGC TGGGGACACA	5340
ATTATTAGAA CTGGCGGAAG AGAAAGCTCG TGAATTTGGT AAAAAAGCAC TGTGGCTAGG	5400

TGTCTGGGAA CACAATGATG CGGCGCGGGC ATTTTATAAA ACGATTGGCT ACCGTTATTA	5460
TAGTCAACAT TCGTTTTTTA TGGGAGACGA TGAACAGACG GATTTTATTT TGTAAAGGA	5520
ATTGTAGCcA AGAAAATTTA GAAACAAAGA TTTCTGATAA AATCAGAAAT CTTTGTTTAT	5580
TTTTTTGACT AAAATTAAGC ATTTTATTAT TTTCAAATG TTATTTTAA AAACGTAAAA	5640
ATTGAAAAAG AAGATTTTGA ATCTGTGAAA AGTTGTGCGAT CTTGTGAGAA AACTCTTTTG	5700
TCAGTGTCTT TTTTCAGGAT GGGTCTCTTC CTGAAAaTCC ACTTCAGTTT ATCTGTTTTG	5760
TGTGTGTTTG ATTAAAAAAT GTTGTGTTGC ATTTCAATTA ACTGTATAAT TtaTTTATGT	5820
CATTTTTAAT AATAAGCGTA TTTTAGCATT TGAmAAGAGA ATTTGTACAT AATATGTGAA	5880
TAAAAAGATT TGTTTCTTAC TTTCTAGAGT GAGAGGACGA TATAAAAAAG GTTGCTATTT	5940
TTTCGTTCAA AGTGAACAAA TAATCGTTTG TTATTTAGTT GGGGTAAC TAAGATTATTG	6000
TTATAAAAAG CGCTTACAAT TCGTGTGATG AGAGTAGTAC GCATGTTTGT GCAAAAGTGA	6060
CAAAAAGAG AAGGCAGGGT GTTTTAATAG AAAAAATTTT TTAATAAAAA GGAGAGTTTG	6120
TCGTGGAGAA AAATGTGTCT CAAGTAACAG CTCAAGAAGA AACAGCGCTG AAAGCACGTG	6180
GCTACAATGA GGATTTGTTG CCAAGTAGCC CGAAACAACG GACGATGGGC GCGCGCAATT	6240
TTTTTACATT ATGGATGGGT TCGATTCATA ATATCCCGAA TTATGCAGCA GTGGCGGGCT	6300
TTATTTTCTT AGGTTTATCG CCGCTGCAAG TTATGCTGGC AGTTGTGTTA AGCTCATTCA	6360
TTGTAGCAAC TTTCATGAAT TTGAATGGCG TGGCTGGCTC TAAATATGGC ATTCCTTTTG	6420
CGATGCACTT ACAGTCTACT TACGGCAGTC TCGGGGCAAA ACTGCCAGGG TTCTTAAGAG	6480
GTTGTGTGGC CGCGATTGCT TGGTTTGGTC TGCAAACATT TACAGGATCG CTGGCTTTAT	6540
TAATTATTCT CGGGAAATTT TGGCCGAAC TTTTAGAAAT TGGTGGATCA TTCCAGTTTT	6600
TTGGTCTCAG GCTACCTGAA TTAATGGCGT TTAATCTTTT CTGGCTCTTA AATGTGGCAA	6660
TAGGTTTCGG TGGTAGCAA ATTTTGAACC GGTTTACCGC CATTTTAAGT CCGTTGATTT	6720
ATGTCGTTAT TATTGGATTA ACGATTTGGG CCATTCGAGC TGGGGGCGGT TTAACACCAA	6780
TTTTGTCCTA TCAAGTTTCT GGTGCGATTC GTTCAGTCAA TCCACTGGTG GCGTATCTGA	6840
TTATTTTAA CTCGGTTGTG GCCGTTTGGT CAGCGCCTGG TGCGAGTGTG GCGGATTTTA	6900
CTAAGAATGC GCGCTCGACA CGTGCTCAGG TTGTGGGGCA AACAACAGGT TTAGTTGTTG	6960
GTTATGGGAT TTTTCGTTTT TCCAGTGTGG TTATTCTGTT AGGTGGTTCG CTGTATTTTG	7020
GGATTCAAGA ATGGAATATT TTAAACATTA TTGATCGTTT AGATAATGTG GCAGTGGTGG	7080
TCTTAGCCAT GTCGGTTTTT TTATTAACGA CTATTTCAAC GAATGCAACG GGGAACATCA	7140
TTCCGGCAGG TTATCAATTG GCGGCGCTAT TTCCTAAAAA AATGACGTAT AAAAAAGCGG	7200
TCATGATTGC CAGTGTATC AGCTTTCTCA TTATGCCTTG GAAACTGATG GAAAACGCCG	7260
ATAGTATTTT TATTTTCTG AACGCAATTG GGGCTGTTTT AGGACCAGTC GCTGGTGTGA	7320
TGATTGCGAA TTATTATTTT GTTCAAAGC AACAAATTGA TTAAATGCT TTATACGTAG	7380

ATAAACACAA AAAAGAGGAA GCCAATCCTT TTTACGGATT AAACAAGCCG GCGTATGTTG	7440
CTACTATTCT GCGGTTGGTT CTTTCCTTAA GTGGTCAATT TATTCCACAG GTTAAAAATCA	7500
TTGCTGATAT TTCGTGGTTT GTGGGCTTTG CCACAGGTTT TGTGCTGTAC TTAGTGTTAA	7560
AAAAATGGAC TTGGGATTCA AAAAAAGTGA AAGAAACAGC TTATCAGGAG GGAAAAATAAA	7620
TGCAAGCAGA GTTAGTGATT AAAAACGGGT TAGTTATTTT AGAAACTGGA GAAGTCATTA	7680
CGGATGTCGC CGTTCAAGGA GGAAAAATCG TAGCGATTGG GCAAGACTTG TCAGGAGAAC	7740
GTGTGATTGA TGCGACGGGT TTAGTCGTTA GTCCTGGGAT GGTGGATGCT CATGTTCATA	7800
TTACCGATCC TGGTGGTGGT TACCGCGATG AATGGGAAGG GTATGTCACA GGAACGGCCG	7860
CTTGTGCCAA AGGGGGCGTG ACGACCTTTA TGGAAATGCC TTTAAATCAG ATTCCAGCAA	7920
CTGTGGATAA GACCTCTCTT GAGATCAAAT ACAAGGCAGG AGAAAAATAAA CTAAAAGTAG	7980
ATGTAGGGTC ATTTGGGGGC GTGGTCCCGA CCAACTTGGC TGATGGGATT CAGGAAC TTG	8040
ATGAAGGAGG CGTTTCTGGC TATAAATGCT TTTTGGGAAC GTGCGGCGAT CGTTCAATTG	8100
AAGGAGATTT TCAAAATGTC GACGATTATT CATTATATGA AGGAATGAAA CAAGTCGCCA	8160
AAACTGGCAA AGTGCTAGCA ATCCATGCAG AAAATGCGCC AATTACGGAT AAATTAGGCG	8220
CGGTAGCTTA TCAAAATGGC GAAACGACTT TAGCGGCATA TG TAGCTACA CGTCCTGTGT	8280
TTACCGAAGT TGAAGCGATT CAAAAAGCCA TTTTATTTGC CAAAGAAACA GGTGTGCGGA	8340
TTCATATTTG CCATGTGGCT TGTCAAGAAG GCGTGGAAGA AGTTTTGAAA GCCCAAGCAG	8400
AAGGGGTAGA TGTGACTTGT GAAACATGTA CGCATTATTT ATATTTTACC ACGGATGAAC	8460
TTGATGCCAT TGGTCCAGTT GTCAAATGTT CACCGCCGAT TCGTGACGCG GACCAACAAG	8520
CCGCCTTATG GAACCACGTG CAAACAGGCG GCATCGCGTT TGTAACCTCT GACCATTCTC	8580
CTTGACGCC AGACTTAAAA GATACAACCA ATGCTTTTGA AGCTTGGGGC GGCATTTCTG	8640
GTGTTCAAAA TAACGTCGAT GTGTTATTCG ATGAAGCCGT CCAAAAACGT GGT TTATCCT	8700
TGAAACAATT CGCGGATATG ATTGCAGCAA ACCCAGCTGA CCGTTATCAC TTAGCACAAA	8760
AGGGCCGCAT CAGCATTGGT AAGGATGCCG ATTTTGTACT AATTAAACCT AATGCGCCGT	8820
ATATCTTAAA AGCAGAAGAT TTAGAATACC GAAATAAAAT TAGTCCTTAC ATTGGTCGTG	8880
AAATTGGGGC CCAAGTGATT CAAACAATTT TAAGAGGCGA AACAATTTAC GCACAAGAAA	8940
CTGGGGTAAC CGAGGCGTTT AACGGCGTAT TTATCAAAAA TTAAACACCT ATTTTAGAAA	9000
TGAAAAAGGA CAGAAGCGTG ATTGAATTAG CGCCTCTGTC CACGAGTTAG AGGATGTTTG	9060
TTGGTTGTTT TGCACAAAAA AAGGCTTAAT TTTGTTTGAT AAGTACAATG AAGATGGCTA	9120
GGGAAACGTT TACAATCAGA GTGTAATCAG TCAGGAGGTT TTGTGATGGA CTTGGAAAAA	9180
GTTTTTAAAC AAAGAATCGA AGAGTTGTCA GCAATTGGGA GTGATCCTGC AGGAGGGATG	9240
ACACGCTTAC TTTATACAGA TTCGTGGTTG GCAGCACAAA AGTATGTGCA AAGCCAAACT	9300
GAGGCGTTTG GACTGGAAAC GCAATTTGAC GAAGTCGGG <sub>a</sub> ATCTGTTTTG TCGAGTGGCA	9360

GGAACGGAAT TTCCTCAGGA AACGATTTTA ACTGGGTCAC ATATCGACAC AGTAGTTAAT	9420
GGTGGGACCT TAGATGGTCA ATATGGTGTC ATTGCTTCCA TGACAGCTGT TCAATATTTA	9480
TTAGACAAAT ATGGGAAACC ACGGCGTTCT TTAGAAGTGC TTTCCATGGC TGAAGAAGAA	9540
GGGAGTCGTT TTCCAACGGT TTTCTGGGGA AGCAAAAATG TCGTTGGTGA AGCCAAACGT	9600
GAAGAAGTAG TCGATATCTC TGACAACGAA GGTAAAAAGT TTGTTGATGA GATGCATCGC	9660
CAAGGCTTTG ATTTTCGGCA AGAGCAAACCT AGACGTGAGG ACATTGAAGC ATTTGTCGAA	9720
ATTCATATTG AACAAAGCAA TATTTTAGAG AACGAAAAGC TCCAAGTCGG TGTTGTTTAT	9780
AGTATTGTAG GGCAACGTCG CTACACGGTT AACTTAAAGG GGCAAGCCAA TCATGCAGGT	9840
ACTACGCCAA TGAGTTACCG TCATGATGCG GTGTACGGCT TTGCTAAAAT TTGTGCAGAA	9900
GCCATTGATC GTGCGAACGA AATTGGGGAT CCGTTAGTTT TAACATTCGG CAAAGTCATT	9960
CCGAAACCCA ATACAGTCAA CGTGGTGCCCT GGAGAAGTGG AATTTACAAT TGAATGCCGT	10020
CATACCGACG CTGCTTTTTT ACGACATTTT ACAGGTGAAT TAGAAGAACG GATGAAGACA	10080
ATTGCTCAAG AATTAGGCTT GACGATTGCG ATTGATCGCT GGATGGATGA AGCCCCTGTA	10140
CCGATGAATC AAGCAATTGT AGAAGTTATT GAACAAAAG CCAAAGAAGG CAAGTATCAG	10200
TATCGTATGA TGCATAGTGG TGCTGGGCAT GATTTCGCAA TTATTGCGCC TCATTACCCG	10260
ACAGCGATGA TATTTGTGCC AAGTATTGGT GGGATTAGTC ATAATCCTGC AGAAGCAACG	10320
GCATTTCCAG ATTTGGTTGA AGGTGTCAAG CTTTTAGCAG ATACCTTATA TGAATTAGCG	10380
TATAAATAAG ATAGAAAAGG AAGAATTGTG ATGGGTTATA AAAATAATCG AGTTGGTTAC	10440
CAAAAAGAAT TGTTAACATC AAGAGCTGTG ATAAAAAAG ACAATTACGC CATTATTCCC	10500
CATGATGGTT TGGTGCAAAA TGCAGTACCT GGTTTTGAAA ATGTCGATAT TTCCATTTTG	10560
GGATCACCTA AATTAGGTGC GACATTGTG GATTACATTG CTACGTTTCA TAAAATGGT	10620
CAACAAACAA CTGGTTTTGG TGGCGACGGG ATCCAAACGT TGGTTTATGT CATTGACGGT	10680
CGTTTGCGCG TTTGAGATGG TCAAGAACT CATGAACTG AAGCCGGGGG ctACGCCTAC	10740
TTTACGCCTG AAATGAAAAT GTATTTAGCG AACGCACAGG AAGCGGACAC AGAAGTTTTC	10800
TTATATAAAA AACGCTATCA GCCGTTGGCA GGCCATCAAC CTTATAAAGT GGTGGGTCA	10860
ATTCATGACC AACAAACCGA AGAATATGAA GGAATGACAG ACGTGTTACT TTGGAGTTTA	10920
TTGCCgAAAG AGTTTGA CTGATATGAAT ATGCATATCT TATCTTTTGA ACCAGGTGCT	10980
AGTCATGCCT ACATTGAAAC GCATGTCCAA GAGCATGGCG CGTATTTAAT TAGTGGCCAA	11040
GGGATGTACA ATTTGGACAA CGAATGGTAT CCAGTGGAAA AAGGGGACTA TATTTTCATG	11100
AGTGCCATATG TGCCACAAGC AGCCTATGCG GTTGGTCGTG AAGAACCGCT AATGTACGTT	11160
TATTCCAAAG ATGCCAATCG TGAACCAGAA TTATAAAGAA AAGCCGAATG TAAAGTAATG	11220
ACCTCGGAAC AATAGGATTT GAACACGCAG tGgTGAACGG TAGAGTCCTT CTCTTGGA	11280
GACGTTACTT TTATTCGGCT ATTTTCCTGT TAATTATTTT TCTGAAAGTT CTAACGCTAC	11340



CCGTAAGAGT AAACCTCGTCT CAGGCTCTTG GATAGAGGTG CCTAATAATT TTTCACATTG	11400
ATTCATGCGA TATTTAATGG TGTTTCTGTG AAGGTAAAGA GCGTTGGCCG TTTTAGTAAT	11460
TTGCGAATTG AAATCTAAAA AAACCTTCAA GGTTTTTCGT AATTCTTGTA ACGTCGGTTC	11520
CGTCGGATAG GCCAACTCTT TCAATTGTTG CTGGCAAAAG TATTCACAT CTTCAGTGCC	11580
AATCTTTTCA AAAAGTCCAG CTAGGCCTTT GGGATGGAAC AATTGCACAG TCGCTGGTTT	11640
TTGAGCGTGT AAAGAAGCCT CTAATGTCGA GGAAGCTTCA ATATAAGAAT TAGGCAAATC	11700
TTCTAGGTTT TCATAAGCAT TGCCTAAAGC AAAGCGAATG GTAATAGGCA AAGCCTGTTG	11760
CAAACGTTCA GCGAGGTTTT GCAAAATCAT CAAATGGTCA TTTTTTTTAG ATTGAAAAAT	11820
CAAGATACTT TTATTTTGAT TTTTAAATTT AAACAAGGCG ACATCTGGCA GGATTTCGG	11880
CAATTGTTCT TTGAGCCATT GGAAAATTAG TTGCCCTTCC GCTTGCTGAT AACGAATATG	11940
GGTTTCATTT TCTGGTTTAG TCACGCCGTA GACAATCGCT AATTGATAAT AATCACTGTT	12000
GATTAAACGA TAATTTTCC AAAGTTCCAA CCAATTTTGA TGTTTACTCA GCGCTTCTTG	12060
ATGTGTATCA AGCAAGCGGT CCAAAAAGTC TGTTTGTAGA TGCTCAAAAG ATTCAGCAAT	12120
TTTCTGGTTT TTAAATAACA TAAAAGTCAA AACTAAAATG GCCTGATCAA TGGCAAATTC	12180
AGAAATAGGA TAAGGAATCT GTTCAGGAGA TAAGACCAAT AGATAATAGG GAAAATAATC	12240
GTTACACGA ATTGGAAAGC CAGCTACTTG AATGTTTCT TGTCGCTCAT CCTGTAAACG	12300
AAAGATTTTC TTTTCTGTG CCAATTGTTG AAAGTGGTCT TTTGATAATT GTTCGATGAA	12360
AAACTCCGCC GATTTTGTAT TGCCATAAAA ATAATTCGAG TGGGCAATCA CCTGTTGCCA	12420
AGGACTTAAT AAAATAATTG GCGCATTTAA AATTTTCGCA AATTCAGCGA TGAAGCGCGT	12480
AGCATCAACA TCTTGCTTA ATAATGTTGA AAAACGCTTT TGTATGTCAA ATGCTACAGA	12540
CATCTGTTCTG GTTTTGCTAT CACGAAGATA GCCTACAATT TCATGCAACA ACCCGCCTAG	12600
TGGCTGTGTG CTAGGAATTT TAATCAGGGG CAAATCAACG GCACTGGCAT AGGCAACAAT	12660
TTCTGGTGAA ATTTCTGCAA GGAACCGCCC AACTTTGATT CCTAAAGCCG TACACTCTGC	12720
TTGTTTTAGC GAGTCAATAA AAGGTTTCAA TTTTCTTGA TCATCTTTAT AAATCATTGC	12780
CGTTGTAAA ATCATGACAT TTTTAGGGAT AAAATCAGCC ACATCAGGTG TTTCCGTGAT	12840
TTCTACACTT TCCAAGGGCT GTGTTAAGTT TGAGTGAGAC GATAATAATT GTAAGTCAGA	12900
AAATCTTGGT AAAGCCAAGA CCTCAGCTAA TATTTTCATC ACAACAACCT CCTTTTTTTA	12960
TAGTATAGCG AAGAATAAAA AGGAGTACAA AAGATAGTTA TTGAATATGT CTAAAAACA	13020
AAGTCAAAAT AGTGCAAATT GAACAATGAA TTTCTGGGTT TTTCCAAAGA TAATAAAAGA	13080
GAAGAATTAA AAATGAAGTG AGGGATTACT ATGTATCAAC AACCAAATTT TGTACGGCGA	13140
ACAATCATGA CGCCGGGACC AGTCGAAGCC CATCCATCGG TACTTCGTCA AATGGGTCAA	13200
CCAATTTTAG GGCAATTTGA TCCAGAATTT CTACAGATTA TGGACGAAGT TCGGGAAATG	13260
ATTAAAGTAC CTTTTGCCAC AAAGAATCAA CAAGCCTTTG CCATTGATGG CACGTCACGT	13320

TCAGGCTTGG AAGCGGGGTT ATTTGCGTTG ATTGAGCCAG GCGACAAAGT GCTTGTGCCT	13380
ACGTATGGCC GGTTCGCTTA TTTATTAGGT GAAATTTGCG AACGCGCCCG GGCCGAAGTC	13440
ATTTATTTAG AAAAGGACTG GCTCGCACCG TTTGAACAAA CAACGGTGAT TGAAGCGATT	13500
AAAGAACATC AACCAAAAAT TGTTCGATG GTCCATGGCG AAACGGCGAA TGCACAAATG	13560
CAACCACTGG ATCAAATTGG CGCCTTTTGT CGGGAAAATG AAATTTTCTT TGATAGTAGT	13620
ATGGTGGCGA CTTACGGCGG TGTGAAACA AAGGTCGATG ATTGGAAGAT TGATATTGCG	13680
GTAGCAGGCA CTCAAAAATG TGTCAGCGTA CCTTCAGGCC TTTCTTTAAT TACCTACAAC	13740
CAACGTGTAG CCGATTATCT GCGGGCCGT TACCAAAAAG AGTTAGGGCT TGGCGCGGAT	13800
GCGCGGAATG AACGGTTCAT TCAAAGTAAC TACTTGATT TATCGCAATT AGAAAAATAT	13860
TGGGGACCAG AGCGCTTAAA TCACCATACA GAAGCGACTA CAATGATTTA TGGGCTACAT	13920
GAAGGGCTGC GCTTGTTACT TCAAGAAGGC ATGGAAAATG TCTATGCACG CCATCGGAAA	13980
AACAATCGTA TTTTAGTTGA AAGTCTGCAA AAAATGGGCT TAGAAATTTT CGGGAAGTTA	14040
GACACAAAAA CACCAACAGT CATTCCTGTG GTCATTCCAG AAGGAATTGA TGGTGAAAAA	14100
GTGCGGAGCT TATTACTGGA TCATTTTAAA GTTGAAATTG CCTCTTCCTT CGGTGATTTA	14160
AAAGGAAAAA TTTGGCGGGT CGGTAATATG GGTTACAGCA GTCGCGAAGA CAATGTGTTG	14220
CATTTCTTAA GTGCCTTTGA GACTGTATTG AAACACCAAG GCTATCAATT TGAAGGCGGT	14280
AGCGGTAGTA CTCATGCATT GGCTGAGTAT TTAACTAAA AAAGAAGGTA GGACAAAGGT	14340
CACTACGTAA AAAGCGTAGT GCTTTTGTTT CTTGTGAAAC GTTAACCAAA AAAATGATTG	14400
GTGGAGGAAA AGAAAATGAC AGACAAGACA AAAAGTTCGT TGGCTTATTG GAAAAAATC	14460
GTTATTTTAC TATGTATGGG CTGGGTAACA ATCTGGATTT ATCGTTCCGC GCTTTCTCCT	14520
GTTTATCCAC AAATTCGTGC GTCTCTAAAT GGTAACGCTT CCGATACGGC CCTAGGGTCC	14580
ATTTCTAGCT TTTATTTTCT CGGTTACGTA GTCATGCAGA TTCCTGCAGG CTTCTTGGTA	14640
GACAAAATTG GGAAAAACG TGTATTAATT CCAGGCTTTA TCTTGTTTGC GCTGGCAGCA	14700
TTACTGATTG CCAAGCCCA AACGATTGGA ACAATTTACT TAGGTAGCTT ATTGGCAGGC	14760
TTAGGATGTG GTTCTTTCTA TGGCTCGGCT TACTCGTTGA CGTCGCAAAA CATTCCTCAA	14820
GCCAAAAAAA GTTTTGCAAC AGCTATTGTC AACAGTGGTT CAGCCGTTGG TTCAGGTTTA	14880
GGGATGATTT TATCTAGTTA TTTAGTTGCC CAAAAGGAT TACCTTGGCA AGTGATGATG	14940
TACATTTCCG CCTTTATGAT TATTTGTATG TTAGTGGCTT TTCAAGTGAT TATTCGTAAT	15000
CATAAAGAAG ACATGGCCTT GATTCAGCCA ACCGTTGTTG ATGAAAAAGA TAACGGAGCG	15060
GTAAAAGAAA AAGTGCCCTT TAAAAAATG TTTGCGCCAC ATATGCTATT TGCGTATATT	15120
TTATATTTTG GCACTTGTTA TGCATATTAT ATGATTGTGA CTTGGTTACC TAATTTTTTA	15180
AGTACCGAAC GTGGTTTCCA AGGAGCAGCA ATAGGGTTAT CCTCTTCACT GGTCGCATTT	15240
GCCTCTATTC CAGGAGCGTT ATTTTTTAGT CGGTTGGCGG ATAAGTACAT GCATAAAAAA	15300

GTTCAATTCA TCGTTGTTTT AGAATTTTTA GCAACAGCCA TGCCTTGTT AACTGTACAA	15360
GTAAGTAATG CGACGTTGTT ATTAATTGCT TTAATTATGT ATGSETTCTT AGGAAAATTA	15420
GCTGTAGAAC CGATTATTAT TTCTTGCTA GGTGAAATG CACCAAAAT TGGAAATGGC	15480
ACTACTTTGG GCGTGTAA CTTTTTCGGC ATGATGTCAT CTGTGATTGC CCGAGCTTTA	15540
ACTGGGAAAA TTTCTGACAC AACGGGTCG AAAGTCATGG GCTTTTATAT CGCCGTCGTG	15600
ATTTTACTAA TTGGCAGCAT TCTTTTTATG CTTGTCAATT TAAAAACG CAGGGCAGAT	15660
TCGTCTAATA AATAAAAAA CAACGCAAC TTATCTTTCG CTGATAAGT TTGCGTtGTT	15720
TTAgcTTTT GACAATCATA TTAATCGCTT GTGCTAATCG CTCTTCTTGT TCTTTAGGAT	15780
CTTTTTCAGG ATTTTCAAAG CAATGCTTTA AATTCTCCGC AACTATCATC CCCATTACGC	15840
GATCAATACT AGAACGGACC GCACTCAACT GAGTGATAAC ATCGATACAT TCTTTTTCTT	15900
CATCAATCAT TTTTGAATC CCGCAATTT GaCCTTCTGT GCGGCGTAAA CGATTCAGCA	15960
CTTCTTTTT TTCTTCAGTC ATTTTGCCCC TCCTAAGAG TAGTGATTGG TTTAAAATAT	16020
ACCTACACGA GGTATTTTTG TCAACTATCT TTTCTTGAA TCTAFAAATG GCAGAAATCT	16080
CTTCAGTTGA CAAGtCTAaC TTTTtCATGT tAAATATmCC CCGTAAGgTA TATAAGAGAA	16140
GGAGAAAAaG rTGTTATTTT AATTGTTTAA AAAACGCCA AGTATTTCTrA CAAAGGAATT	16200
GCAGTCACGT TTATCTAAAG AGATAACGTT ATTAGATGTC CGGACACCTA GCGAATACCG	16260
CGCAGGGCAT ATTCCACAAG CGATTAATGT CCCACTTAAC AAAATTCCCG CCTACAACAA	16320
ATCCGCTAAC GAAGTCTATG TGATTGCCA GTCCGGCATG CGCAgCAAAA ATGCGGCCAA	16380
AATTTTAGCG AGAAAAAAT ATCACGTGAT TAATGTTCTG GCGGGATGT CCCAATGGTC	16440
TGGTCAAATT AAAGGAGGAA AATAAATGAA AATTGTAATT ATCGGAGGCG TGGCTGGTGG	16500
TATGTCAGCA GCGACACGTT TACGTCGTTT AATGGAAGAT SCTGAATTG TTGTCATGGA	16560
AAAAGGCCCT TTTGTTTCAT TTGCAAACTG TGGTTGCCT TACTACGTTT CTGGCGAAAT	16620
CGCAGAAAGA GAGCAACTGC TTGTTCAAAC ACCCGAAGCG TTAAGGCAC GGTTTAATTT	16680
AGATGTTCTG CCTCACCATG AAGTCGTGGC TATTGATCCA ATAGAAAAAG TGATCACAGT	16740
GAAACATGAA ACAGAAATTT TAACAGAACA CTATGACAAA TTAATTTTAT CACCAGGGGC	16800
GAAACCTTTT GTGCCACCAA TTACAGGATT GGCAGAAGCC AAAATGTTT TTTCATTAAG	16860
AAATGTTCTT GATTTAGATC AAATTATGAC AGCCTTGACA CCAGAAACAA AACGAGCCGT	16920
CGTGATTGGC GCAGGCTTCA TTGGCTTGA AATGGCAGAA AACTTGCAAA AACGCGGATT	16980
AGAAGTCACT CTCGTGAAA AAGCGCCTCA TGTTTTACCG CCATTAGACG AAGAAATGGC	17040
CGCTTTTGTC AAAGCTGAAT TAAGCAAAAA CAATGTTCAA GTPATTACGG GACAATCTGC	17100
GTTTGCTTTT GAAGAAGAAG GGCAAGTGAT TCGCTTAGAA GACGGTCAAA CATTAGCTTC	17160
TGATTTAACC ATTTTGTCGG TGGGGGTCCA ACCAGAAAAT ACCTTAGCAG TTGAAGCAGG	17220
TGTAGCAACT GGTTTACGTG GCGGTATTGT TGTTGATGAA CACTATCAAA CGAATCAACC	17280

CGATATTTAT GCGGTTGGGG ATGCTATTGT AGTGAAACAA CAAATCACTC AAGAAGATGC	17340
GCTGATTCTTCT TTAGCTTCTC CCGCCAATCG CCAAGGACGT CAAGTAGCGG ATGTGATTGC	17400
TGGGTTAGAG AGAAAAATC AAGGAAGCAT TGGGACCGCC ATTGTGCGAG TCTTTGATTT	17460
AACCGCTGCT TCAACTGGTT TAAGCGAACG GGCTGCTAAA GCTGCTGGAC TAACAACAGC	17520
TGTTGTGCAT ATCAGTGGAA AAGACCATGC GGGGTATTAT CCTGGCGCAA CAGATCTTCA	17580
GTTAAATTA GTTTTCCATC CTACGACAGG GGAAATTTAT GGCGCACAAG GAATTGGGGC	17640
AAAGGGCGTA GATAAGCGGA TTGATATTCT TGCACCGCT ATTAAAGGAC AGTTAACTAT	17700
TTTTGATTG CCTGAATTAG AGTTTACCTA TGCGCCGCCG TTTGGTTCAG CGAAAGATCC	17760
TGTGAACATG TTAGGCTATG CAGCGATGAA CCTTGTAGAA GGATTGAGTG AAAACGTTCA	17820
ATGGTATGAG TTATCCAACG AATTAGCTAA TGGAGCTGTT TTATTAGATG TCCGTAATCC	17880
CGCCGAACGA GCCAATGGTC AATTTAAAAA TGCTGTGTCT ATTCCTTTAA ATGAGTTAAG	17940
AGAACGTTTG GAGGAATTAG ACAAGTCAAC GGAGTACATT GTTAGTTGTC ACAGTGGTTT	18000
GCGTAGTTAT ATTGCAGAAC GGATGCTAAA ACAAGCAGGC ATCTCAGCCA AAAATTTAGA	18060
TGGTGCTTTT GCGCTATATC GAATGGTAAA ACCGGAGGAA CTAGAAAATG TATAACACGA	18120
TTGGCATGCC TGAATTTTAT CaGGAAGCAA AACGAAAAAA ATTACCGATT ATTGATGTAC	18180
GAGAAGAAAT GGAATACCAA CTGGGGCATG TCCCTCAGGC CATCAACTTA CCTTTAAGTA	18240
CTTTAACCGA AAATTATCAG CAATTAGATA AAAACCAACC CTATTATCTG ATTTGTCAAA	18300
TGGGGTCGCG CTCCGCACAA GCGTGTGCGT TTCTCAGTCA ACAAGGGTAT CAAGTAACCA	18360
ATGTTCTAGG TGGTACATCT GGCTGGTTGG GTCAATTAGA ACAATAAAAT AACAAAAAC	18420
GTTTATTCTT TTACACAGGA TAAACGTTTT TTTGTTAGAT AAAGATGTGT CAAAACGTGA	18480
AGAGGAATGA TAGGAAGATG AAGGGACCAG TTTTTTTATT CATTATATAC TAAATAAAAT	18540
GACAACTCG TCAACAGAAA TCAAGGAGGA ACAAATAAAT GAAAAAGTCA GGCCAGAAGA	18600
AGCGAAATAA AAAAGTGTGG TTCGGTATTG GTGCCGCAGT GGTGTCGTC GGATTTATCG	18660
GAGCAAAAC AGTTTTTTCT TCAAAAGAAG TTGAACCTGA ATACACAACG TATACCATTA	18720
CAGAAATGGC CTCATTAAAA TTAGATGGTC AAGTGAGTTT TCTAGATACT CGGGATATTT	18780
TCTTTGATCC TTCTTTAGGA AAAATTGCCG AAATAAATGT TGAAAATGGT AAAGAAGTCA	18840
AGAAAGATAG TCCATTGTTA ACGTACAATA ACTCAGACAT CCAAGCCACA GAAACAGAAC	18900
AAGCGAATGC TGTTAATCGG AACATCTTC AAGTTCAACA AGCACAAGAG AACGTTAAC	18960
TAGCCACTCA AAAATACAAT GAAGCGTTAA ACAAAGTTGC TGCAGCCAAG CAAAAATTAA	19020
ACACCGCAAA AGAGGCTGAA GAGAAAGAAA CACTAAATGC GGAAATTCAG CAATTGAATG	19080
AAGCCGTTAG TGCTGCGAAT AGTGAAGTCG CCAAGCTAA CCAAGCCCTT CAATTAGCCA	19140
ACAGTGATGC CGTCGGAGCA GCAAACACGC TTGAACAAAC ACGTGGCAAA GTCAACACAG	19200
TCGTTACAGC ACCAATTGAT GGCCAAGTCA CTGTTGATGC TAGTGCCATG AGTAGTACCG	19260

ACAAACCAGT CATTAAAATT GCAACACAAA AGAAAAATAT TCAAGGAAAA GTAACCGAGT	19320
ATGATTACGA TAAATTACAG ACAGGGGAAG AAGTAACAGT GACGACCGTT GGTTCAGGTA	19380
AATCAGCACC AGGCAAAATT GTTTCATTG CGCAAACCCC AATAGCTAAA AATGAAGGCA	19440
ATCCAGTAGT GAGTTATCAA TTTACAGTAG AAGGTGATTT TCCTTGGGCA GAAGGCTTGT	19500
CTACTAGTAT CGCTGTGCCA CAAAAGCAAA TGATTATACC GACAGCAGCA ATCCAAAAAG	19560
AAGGCCAAAA AGAATTTTTC TATGTTTACA AAGCGGGCAA AGCGAAAAAA ACACCGATTG	19620
AAACAGAAAC AAATTTAGGT CGTAAAGTTG TCAAAAGTGG CTTGAATTGG AAAGACCAAG	19680
TGATTGCCAA TCCCAATAAA GAATTAAGA ATAATCAAGA TGTCAGGTA GCTGCCAATG	19740
ATTAAGCTAG CCCATATTCA TAAATATTAT TATTCCGAAG AAGAAACACT GCATGTGTTA	19800
GATGATATTA ACCTCCAAGT AGACGCTGGA GAATTTTTAG CGATTATGGG TCCTTCTGGT	19860
TCTGGGAAAT CGACTTTGAT TAATTTGCTA GGCTTTATTG ATAAAAAGTT TGAAGGAACG	19920
TATTTATTTG AAGACCGTGA AATTGGCGAC TTTTCTGATA AGGAACTATC CCGAATTAGA	19980
AATGAAGCAG TCGGCTTTGT CTTTCAGAAT TTTAGTTTAA TTGAAACACT AACAGTTGAA	20040
GAAAACATTG aATTACCTCT TTTATATAGT GGGTTAACCC CAAAAGAAGC CAAAGATCGT	20100
GTCCACGAAG TTCTGACAAA AGTCGGCTTG CCAGATAAAG GCAAGAAGCA TCCAAAACAA	20160
CTTTCGGGCG GACAACAACA ACGTGTGACC ATTGCGCGTG CGATTGTCAA TCGGCCAAGT	20220
TTTATTATTG CAGATGAACC AACTGGGGCT TTAGATAGTA AAACATCGGA AGAAATTTTA	20280
ACGCTCTTTC AGCAATTAAA CAATGAAGGT GTCACCATTA TTTTAGTGAC ACATGATGAA	20340
GAAACAATTG AATACTGCAA TCGTTTGATT AAAGTTCGTG ATGGAAAGAT TTTAGAGGAG	20400
GTGCTGACAT GAAACGTTCA ATTATTTGGA AAACCGCCTT CCGTTCCATT TTGAAAAATA	20460
AGCGACGCAG CTTGTTAACG ATGCTAGGGA TTGTCATCGG GATCGCTTCG GTGATTACCA	20520
TCGTGGCTAT CGGTAACGGC TTAAAGAAG ACATGGTCGA TAAATTATCT GCTGAAAAAC	20580
AAAAAGAAAA TGTGAAAAAG ATTTCTTTTT CGGCCTACAA CACTTCGGAT ATGTTTAGTG	20640
ACCAAGCGAT GTTTACCGAC AATGATTAG GGGTCGTGCG CATGGTTCCT GGCGTAGAAA	20700
AAATTGATTT TGATCAAAGA GAAGTTGATG GCGCCCAAAA AGGTACACTG AATTTTCAAG	20760
CAGGACCGAA AAATCTTTCT GTAACCTATG AACTTGCCAA CAAAACCACC AAAGAATTGT	20820
TGGCGGGACG CCAACTGAAT GATGGAGACA ATGCAAACT GGATAAAACA GTGGTGGTGG	20880
ATGAACTGGT TGCCAAAGAA CTCTACCAAG ACTCGACTGC AGCAGTTAAT CGAACTTTTC	20940
CATATAAAGA GCAACTTTTT ACCATTGTTG GCGTGACAAC CAATACCAGC GGTGCCATTG	21000
GTCCAGGTAA TGATGACTCA TTGCTTTATT TTCCCAAAA GACCTATGAA CATTATTTTCG	21060
GCAAGCTAAA AGATACATCT ACGTTGAAAC TAACAGTAGC ACCTGGCTAT CAACCAGATC	21120
AAGTATTGAA AGAAACAATA AAAACTCTCT CTCAACAAGG AACCATGAAA AACAGTGGGA	21180
CGTATCAAGA ATATAATGTT AAAGATACCA TCAAAGAAAT GGGCTCTTTA TTAAATAATT	21240

TAACCTTGTT	TATTTCCGCA	ATTGCGGCTA	TTTCACTGGT	AAATTGCTGGG	GTCGGCGTTA	21300
TGAATATGAT	GTACATTTTCG	GTCTCAGAAC	GGATTAAGGA	AATTGGTATC	CGACGTGCAT	21360
TGGGCGGGAC	AGCCAGTGAT	ATTAAAAAGC	AATTCTTAAC	AGAAGGGATT	GCCTTGACCT	21420
TAATCGGCGG	AATTACTGGT	TATCTTTTAG	GAATGATTAT	TGCCTACTTA	GCCTCAATGG	21480
CACTACCGTT	TTCTGTTCGA	CCAGATTTAA	TGACCGTTTC	ATTAGCCATT	GGGATTTCTG	21540
TCTTTATCGG	GGTGTATTT	AGCTATTTCC	CTGCCTCGGC	AGCTTCAAAA	AAAGACTTGA	21600
TCGATATTAT	GAAATAAATA	TAAACAGCAA	AGAAACAGCC	ATTTTTGGCT	GTTTTTTTTA	21660
GTGAAAGTAG	AAAAAAATCC	AAAGATTTC	AGAGAAAGTA	GCTAACAAAC	GTCAAAGGT	21720
TGATATTAGC	TACCAACCTT	TTGACGTTTG	TCAGTTAAAT	AAAAAAGAAA	AAAATTAAAA	21780
AAATAGCACG	TTGAGGATTA	AATAACAAAC	ATTTTTTAAA	TAAGACTATT	GGAAATCCTT	21840
TATAAATCGG	GGATTACAGG	AATGACTGTT	AATGACTGAT	TTGCTTGTA	ATTTCTAAAT	21900
TTTAATTCGA	AAAATTTCCA	GTATACTAAA	TATATCAAAA	CAAAACACGA	AAATTGAGAA	21960
AATTTTGAAC	CAAAAAAGAA	TTAAGGAGGA	ATCAAGAATG	AAAATAATGG	TCATCACAAA	22020
AAATATTTTA	TCAGAACTG	CTTTTCAAAA	CCGCTTACAG	CATTTAGATC	ACCAAGTTTT	22080
TTGTACGACG	TGTCTGATTG	GTGAAATTAA	GGAAAGTCAT	TTGTTTTATC	GAAAATTGTA	22140
TGCCATCATT	CTCAGCGAAA	CGATTCCTCG	CAAAGAACAA	TATTTATATC	TAAAGGAATT	22200
AGCCAGAGTT	AAAACGCCGT	TGTTTCTTCG	CTCGGAAGAC	ATTCCTACCT	CGGAAGAATT	22260
GAGCATTAGC	TATCAATATG	GCATTCTCAA	TTGGTTACGA	TTGGATGCGA	CTTTGGATAC	22320
GTTAAGAGAA	AATCTCCTTT	TGGCAGAATT	AAATAAACCC	ATCCAACGAG	AAGATGAAAA	22380
GTTAGATGTA	GAGGGAATAA	AGCTAACAGC	CGCAGAACGT	GCACTACTGG	CTGTCTTGTA	22440
TAGTACGCCA	AAAGAACACT	TACCAATGGA	GCGCCTGTGT	AAATTAATAT	ATGACACAGA	22500
AGAAACAGCC	AATAATCGAA	AAAAAATGAC	CAAACAGCTG	ACAAAATTAA	ATAAAAAAAT	22560
TCAAGTAGGG	GGCTTTAAAA	AACAAGGTAT	CCAAAATACA	CGGTATAAAG	GCTACAAACT	22620
GCATCCGCAg	TTGCTTCCTT	ACATTGGCGA	ATTATTTACT	TAAATTAGAG	AAGCACTTGA	22680
CGAATGAAAA	TCACCATCAC	AGGATGACCT	CTAAAAACCA	TGCTATACTA	ATACGGAATA	22740
GTATAGGAGG	TGGGTCTGTA	TGAAAGACGG	ACTATATTGG	GTACAACAAT	TAAAAGAACA	22800
GAAAGTAAGT	CGCATAGAAC	TCTTAAAGGA	TATTGAGAAA	AAAGTCCAAA	AGTTAAATCC	22860
AAAGATCAAT	GGCTTTGTAA	CCTTTGATAG	CCAAGAAGCA	GAAGCAACTT	ATCAAAAAAA	22920
TAAACAGCAG	GACACACTAT	TTGCTGGTTT	GCCATTTCCA	TTGAAAATGT	TAGGCCAAGA	22980
AAAAGCGGGG	TGGTCAGCAA	CGGCCGGCTC	GCAACTTTTT	AAAAATAGTA	CTGCAACAAC	23040
GACCTCTAAC	TTTGTCAAAC	AAGCAGAAAC	GATTGGCTTG	ATGCCTTTAG	GACAAACCAA	23100
CGCTCCAGAG	TTTGGCTTTA	AAAATATTAC	GGACCCTGCG	CTTTACGGGC	CAGCCCGCAA	23160
TCCTTGGAAC	CTGGAACATT	CGCCAGGCGG	TTCCAGTGCG	GGCGCTGCCG	CAGTTGTCGC	23220

TAGCGGGATT GTTCCTATAG CCGGCGCTAC CGATGGCGGT GGCTCCATCC GAATTCCCGC	23280
CTCTTTTAGT GGCTTAATTG GCTTAAAACC TAGCCGTGGC TCCATGCCTG TCGGCCCAGA	23340
AGGCTGGCGA GGCTGGCAAG GTGCCTCGAT TGACTTTGCG TTGACTATTT CTATGAGAGA	23400
CACCAAAGCG TTATTTTATG GCCTACGTGG TAGCCACAGT GGCGCCCCCT ACCAAGCGCC	23460
CTTGGCAGAA TGGCAAACAC ACCCTAAAAA ACAACGGCTA AAAATTGCTC TGTGTAcCGc	23520
TTCgcCTATC GGCTCACTGA TTTCGCCAGA AGCAACACAA GCGGCGAAAC AAGCGGCTGA	23580
TTTTCTGGCA GCCGCTGGCC ATGAAATCAT TGAGATTCCCT TATCCAGTCG ATGGTGCAGC	23640
GCTGATTTCG TCCTATTATC AAATGAACGG CGCAGAACT ACAGCGATGA TGAACAGCAT	23700
TCAACAGGGG CTCGGGCGCC CAATTCGTAA AGAAGAAATC GAATCTTTTT CATGGACCAT	23760
GCATCAATTT GGCCAAAAGA TTCCAGCAGC CACCTATGTT CATTCATTAC AGTTGTGGGA	23820
TCAAGCAGCA GTAACGATGG AAGAACTTTT TCAAAAATTT GATCTGTTTT TATCACCAAC	23880
AACGGCTTTT TCAGCACCAA AAATTAACGA AGATTTGCAA AGTGAGCACA TTCGTCAACG	23940
AATGGCGCAA GCAGCAGAAT TAACAGAAGC AGAATTAGCT GAATTAATCT ATGATTATTT	24000
TGATAAAAGT TTACAGCTAA CACCGTATAC TCAACTAGCT AACTTGACAG GACAACCAGC	24060
CATCAGCTTG CCCACACATG TTACGGCAAC GGGGTTACCC TTAGGCATTC AATTGCTGGC	24120
TGCTCGTGGA CGTGAAGATC TCTTATTCCA AGTAGGCGAA CAGTTTGAGC AAGAAGGCAA	24180
ATTTAAATTA CCAGAAAGTT ACCGCTAAAT AGGGTGTTAT TGTGACCAA TTTCAGATAA	24240
AATAGACACA AGACAAAAA CTGAGGTGGA AAAATGCTGT ACGTAGTGAG ACATGGTGAA	24300
ACAGATTATA ATGTGGCCCG TAGAATTTGT GGTCATGCGG AAGCGCAATT GACCGAAAAA	24360
GGCTATCAAC AAGCGGAATT GGTAGCTGAA AAAATTGCCA AACAGGAAT TCAAATCGAC	24420
CGTTTACTGG CGTCGCCTTT AAAAAGGGCG CAAGAAACAG CCCGTAAAAT TGCAGAACGA	24480
AATCAATTGA CGATTGAAAC AGAACCTCGT TTAATTGAGA TGAATTTTGG CATTTATGAT	24540
GGTGAACCGA TTGAAACAAC GGCTTTCCAA GAAAATCGTT CTGAGATTTC CTTGCCGTTT	24600
CCAGAAGGGG AATCAGTCCT AGATGTCGCG GGTGCAATTT ATCCCCTGAT TGAGGAAGCC	24660
TTGGCTTCAG AGGAAACTTA CCTATTTGTT TGCCACAATG CGGTCATGCG AGTCATTGAT	24720
AACTACTTCA ATGGTAAAAA AATCCAAGAT TTCTTGACT TTCATTGTGA AAATACACAA	24780
CTTGTAACAT ACGGAGAATA AAAATCAAGT CTTTAATCTT TTGGTTAAAG ACTTGATTTT	24840
TTTTATGGTA AATAAGAAAC CGCTTCCGTT TTAATGTATA ATAAATTAAC TTTTGGATCA	24900
AAAAAGAAAT GGGGAAAAGA CTATGAGAAT TGATAATATT CATCAAGCGT TAAAACAGTT	24960
TCACGTACAA AATCCAAGTA GCAAAGAAAC CTTGTTGCCC GCAACGGcTT CGCAATTAAG	25020
TCAGGCCTCT GGCTATACCC GAACCGCTGT TAGTGAAGAG CTTTCAGAAT TAGTGAGAAA	25080
AGACCAAGTC GTTAAAGTTA AATCAAGACC GGTGTTATTT TTTGATAAAG ACTACTTAGA	25140
GGAACAGTTT CCTTTTTTTA AAAGTAGCAC GTGCTATGAT TCCCTCGcTT CGCTGCAAGA	25200

GGACATTTAT CAACAGACAG AACAGATACA TAAGAAAAAG CAGCTACTGA AAATGAATCC	25260
ATTTGCGAAC GTAATTGGTT TTGATGGGAG CTTAGCAGAC GCCATCAAGA AGTTGAAAGC	25320
CGCTATTTTA TATCCGCCTA ATGGCTTGAA TATTTTACTT ACTGGAGAAT CTGGGGTGGG	25380
GAAAACCTTA ATTGCAGAAC AATTACACCA ATTTTATCAA GTTAAATGA ACCAAGAAGT	25440
TCCGTTTATT TATTTTAATT GTGCAGAATA TTTCAACAAT CCAGAATTAT TAACTTCGCA	25500
TCTGTTTGGC TATAAAAAAG GCAGTTTCAC TGGTGCCGTC AATGACCAA AAGGTTTAGT	25560
GGAGTTGGCA GATGGCGGTT TTTTATTTTT AGATGAAGTC CATCGATTAA CGAGCGAAGG	25620
CCAAGAAAAA TTGTTTACAG TTTTGGATAA AGGGTATTTT ACACGGATGG GCGArCgAAA	25680
AACAAAGAAA TGTATCGATA AAATTTATTT TTGCAACAAC CGAAGCAGTC AATCAACATT	25740
TTTTGAAAAC GTTTTTAAGA AGGATCCCTG TGACATTAGC CATTCCAGCT TTAGCCGAGC	25800
GCTCTTTGAA AGAAAAAATT CAATTGATTT TGTCTTTTTT CTTAGAAGAA AGTCAACGGA	25860
TTCAAAAGAA TATCCATATT CGTTATGAGG CGTTGAAACA ATTAACATTT ACTCATTTTG	25920
AAGCCAACGT AGGCGAATTG AAAGCGGCGA TTCAATTTAT GAGCGCACAA GCCTATTTAG	25980
AGGTGTTCGA TCAATTAACG GAAACTATTT GCATTGAAGC AAACGAGGGG ACAGTCAAAG	26040
AGCAACCTGA TGTAGAAAGT CTGCTGGCTT TTAATGAGTG GATACCTAAC GATGGCCTCA	26100
CGATTCAATA CAACCATTTT TATTTAAGCA GTCAAGAGTT ACTTAAATCA GAAAGTCTGA	26160
AAAATGATCG TTTTATAAT TTTCTATTGA AAGAATATGC GGATTTGAAA AAATTAACTC	26220
TTTCGACAGA GGATGCGCTA GCAATCATGC AAAACAAAGT CGAGCAACTT TATGATATGG	26280
ATATTTTTGC TCAAGAACCG CTAGCCATAG AGGCCATTGA CGAAACATTG CAGCCAAAAA	26340
TTCAAGAAGC CATTGTCTTT TTTGAACAAT TAATCCAAGA AAAATTGACA GAGGATTTCC	26400
AGAATATTTT AATTGTCCAC CTTTATTCTG CTGTTTTTTA TTCTGCACAA ATGAATGAAA	26460
CATTTTATTC AGATACCAA AATATCTATT CTGGCAAAGT CACCCAGTAT CAAGCTGCTA	26520
AGACCATGGT GGACTTTCTA CAAACATTGT TTGGGATTTT TTTTGCTGAT ACGGAAATCC	26580
TCTTTTTTCG TCTATTTATT CGTAAAGTCA CAGCTAAAAA AAGAGCGCAC TCAGATGAAG	26640
AAAAGTGTGG CGTTATCGTC GTTGCACATG GTGCCGCGAC GGCCAGTAGC ATGGTTGAGT	26700
ACACCAACCT TCTGTTTTCA ACAACGGTCA TGCGTGCAGT TGATATGCCG ATTAATCAAT	26760
CAGTGGAAGA AACATTAGAA AAAGTCCGTA CAATTGTAAA GCAAAATCAG TATAAAAAA	26820
TTATTTTAAT GGTGGACATC GGTTCCTGG TTTACTTTGG TAATGCTATC AGTCAAAAGT	26880
TTCAAAGGGA AGTGCTGCTA ATTAGCAATA TCAATCTATT AACATTGCTA GAAGTGGCCC	26940
GAGAAGTCAT TTATGAATCG ACGGCGTTTG AGTATCTATT ACCAGTCTTA AAAGAAAAA	27000
ATCATCGCGC CTTGATTTGC AGTAAAGGTC GCTTCAACGA GGCAAAAGTT CTCATTATTT	27060
CCTGTCTAAC TGGCATGGGG ACTGCTATTA AAATTGAACA GTTGTTACTA AAAACTTTTC	27120
CCAATGAATT ACTAACGAAT GTACGCATCA TTACTTTAGA GAAAAGAGAA GTGGAAGACA	27180



TTGATCGCTT GCATTCTTTT GTTCGGAAAG AAGAGAAAAT CGTTGGTATC ATTGGCACCG	27240
TAAAGACAGA AGTGCCAGAT ATTCCGTTTA TTTCTTTACA AGATTTACTA TCGAAGCGGG	27300
GCATTCAGCT AGTTTTCGAA TTATTAGGCC ATCATTTAGA AGAAAAACAA AACCAAGAGT	27360
TACTAAAAAA AGTTTCAGCT AAATATATTC AAGGATTATC AATCGAAGCT ATTACAGACT	27420
TACTAACGGT TTTGAATCCG ACAACCATTT CAATCGAATT AGCTAAGATT TATTCAGACA	27480
TTTGCTTGCA AACTAAAATA AAAAGTGACG AAAAGATTTT ACTTCGTTTT ATCATTCAAT	27540
GTGCTTGAT GTTAGAACGG CAATTGTTGA ATCCAGAATA TGACCCGACA GCTTATGaAG	27600
TGTCaTATCG AGAGTTGCCA GAAGCTGTAT CTGTCATAAA AATGGCGTTT CGACCGCTTG	27660
AAGTGTCCTA TAATCTGTTG ATTCCACCTT TGGAAGTGCG CTATATTTAT GAATTGCTCT	27720
TTGAAAAGGC TTAATTCAAA CGGAAATGAA AGTTGGCACA CTTATTGCTT TTAATAAAAT	27780
GATTAATTAG CAAAGGAGTT TGTTTATGGA AAATGAACCA ATTTTACTTT TAACACACAA	27840
TGGTTGGGGC TTGGAATTAG TCAAAAGCGT CCAGATGATT GTTGGAGAAG TCACAAATGT	27900
ACATGAAGTT GCATTACAAG CAGAGGATTC ACTGGGGGAT TATTTAGAAC GAGTCACAGA	27960
AAAAATTGAC CAATTAACGT GGCACAACCA GTTGCTAATT TTAACAGATG TTAAAGGCGG	28020
CACACCTAGT AACGTTGCAT TACGTTTATC AAAAGATTAT GACCTCCTGA TTGTGTCAGG	28080
ATTATGCACA TCGTTATTAT TAGAATCCGT CATGAAACAA AGTGGTCCTG GATTTTCGTTT	28140
AGAAGATGCA GAGATGATTC AACAAGCAGC AGTAGACAGT TGTCAAATTT TAGAAATTCC	28200
AAAATAAGAA AGTACAGGTG AACAGAATGT CAAAAATCGT ATTAACAAGA GTCGATAGCC	28260
GTTTAATCCA TGGACAAGTA GTGACTAAAT GGTTACAACA ATCAGGTGCA AATGAAATTT	28320
TGGTAGTGAG TGATGAAC TAAGCAAGATG AATTTTTTACA AAGTATTTAT TTAATGGCGG	28380
CGCCTCCAGG GGTTCCTGTA GTAATTAAAG GGATTGAAAG TGCCAAAGAA TATTGGGAAT	28440
CGCAAGAAAA GGAAGAGAAA AAGGTGTTAT TTTTAGTGCC AGATTTAACG ACCTTAAAG	28500
CGATGGTGA CCAAGAAATT ATTAAAGAGG AGATTCAAGT TGGTGGCTTA GGTGGCGGAC	28560
CAAATCGGAA AAATGTTTTA AAAAATATCA ATTTATCTGA AGAAGATGTT GCAATTTTGA	28620
CTGAACTTTT AACGAAGGGC ATCCATGTTT TCTTCCAAGC GATTCCCGAA GAAAATCCCC	28680
TACCAATTCA AAAATTAATT GAAAAATATC AATCACTATA ATCAGAGAAA AAGGAGAGAA	28740
TCATATGGGC GTTTTAGGTG TTTCAATTAT TATGGGACTT TATTATTGGT TTGCACGTTT	28800
GCGTTTTGGG TACACATTTT CAGGCATGCT GTCTCAGCCA TTAGGCGCAg cGTTGGTGCG	28860
TGGTTTAGCA ACAGGAAAAC TTAGTGAAGC AATGGTTGTG GGGGCGGGGA TTCAACTTGT	28920
TTATCTCGGA GTCACCTCCA CACCGGGCGG CAATGTGCCT AGTGATCCGG CTTTAGCTTC	28980
AGCAATCGCG ATTCCGATTG CTTTAGGTGT GGAATGAAT GCCGAGGCGG CTGTTGCACT	29040
AGCTGTTCCCT TTTGGTGTGT TAGGCGTCTT TATGGATCAA TTACGGCGCA CCATCAATGC	29100
GACGTGGGTG CATATGGCTG ATAAATATGC GGAAGAGTTA AATTATGGGG GGATTTACCG	29160

AGCTGCGTTT	GTTTATCCAG	CTTTAGCGGG	ATTTTAAATT	CGCTTCCGT	TGGTTTTGTC	29220
GGCGAACTTT	TTTGGTCAAT	CTGTGGTGAA	TAAAGTCTTA	GAAGTTATCC	CAGCATGGTT	29280
GATGCACGGC	TTTGAAGTAA	TGGGTGGTAT	TTTACCAGCG	TTAGGCTTCG	CCTTAACAAT	29340
TATGGTAATT	GGCAAGAAAA	ATTTGATTCC	TTATTTTTTA	ATTGGTTTTG	TTGCAGTAGT	29400
ATATTTTGGC	GCAGAAGTTA	TGGCCGTCGC	TATTTTAGGA	ATCTGTATTG	CCTTTTGTAGT	29460
TAGAAACAAG	GCATTGAGTG	AGGAGGCAGC	ATAATGGAAC	AAATGACACA	AGCAAATGAA	29520
CAACAATTGG	AACAGAAAAA	AATTACTAAA	AAAGAATTAT	CCAAAGCCTT	TTGGATTAT	29580
CAACTAGGCT	GTGAATTATC	AAATTCTTAT	GAACGTCTCC	AAAGTTTAGT	TTTTTGTGCA	29640
TCGATGATTC	CTGCAATTAA	AAAACTTTAC	GCAGACGATG	AAGAACAACA	AAGAGAAGCT	29700
CTGAAACGTC	ATTTGAATTT	TTTAAATACT	GAAGGAACAG	TAGGGGCTTC	CATTCAAGGG	29760
ATTGCAATTG	CTATGGAGGA	AGAAAAATCC	AATGGCGCAG	cGATCTCTGA	TACGGCGATC	29820
ACATCGATTA	AAACTGGGCT	GATGGGTCCA	CTGGCTGGGA	TTGGCGATTC	GATTATTTGG	29880
GCAGCATTAA	TGCCGTTAAT	CATTTCATT	TTTATTCCAA	TGGCTAAAGG	AGGTAACGTG	29940
ATTGGAAGTA	TCGGTCCGTT	GGTCTGTAC	ACAGCCATTA	CATTATACAT	CAGTTGGACG	30000
TTAGTGAATA	AGTCCATATC	ATTAGGGCGT	AATTCAATCT	TAAGTTTGTT	AAAAGACGGC	30060
AAAATTAAGC	AAGTCATTTA	TTCAGCAAAT	GTTTTAGGAA	TGATGATGAT	GGGCGCTTTA	30120
AGTGCGAGTT	ATGTTAAAAA	TGCTAGCCCG	ATGACGTTTA	AAGTTACTGG	CGGTGCCACG	30180
ATTGTTTTAC	AAGATATCCT	AGATCAAATT	ATGAAAGGGT	TACTTCCTTT	AGCAGCAGTC	30240
ATGGCTATTT	ATTTCTTCAT	GGTAAAAAAA	GGCCACGTT	ACGGTATCAT	TATCGGCACG	30300
ATTGTCTTGG	TTAGTTTGGT	CACTTCATT	TTTGGTTTGT	TATAAAGGAG	AAAATAATGA	30360
ATAGTTACGA	AAAATTCCAC	TTAAAAGAAG	TCATTAATGC	TTCTGGCAAA	ATGACCATT	30420
TAGGTGTCTC	CAAAGTTTCG	GAGGCAGTCC	TAGCAGCTCA	ACGATTCCGC	GGCGAACATT	30480
TCTTTGAAAT	GAGTGAACTT	AGTGTGCAAA	CAGGGGCCTT	TTTAGCCAAC	TTGTTGAAGG	30540
TAGAAGATGC	CCAAATCGTT	TCCTCGGCTT	CGGCGGGGAT	TGCCCAATCG	GTGGCTGCCT	30600
TGATTGGAAG	AGGCAGCTTG	TACCACGCGT	ATCATCCCTA	TACAGAAAAA	ATCGAACAGC	30660
GGGAGATTGT	TTTACCTAAA	GGGCACAACG	TGGACTACGG	CACGCCTGTA	GAAGTCATGG	30720
TAGCGCAAGG	TGGCGGCCAA	GTGGTGGAAG	CTGgcTACGC	CAACATGTGC	AGTCCAGAAC	30780
ACGTGGAAAT	GATGATTTCT	GAAAAACAG	CGGCCATTCT	ATACATCAAA	AGCCACCATA	30840
CCGTGCAAAA	AAGTATGCTG	ACGGTGGCCG	AAGCGGCCAA	AGTGGCGCAA	CGCCACAAGG	30900
TGCCGCTAAT	TGTTGACGCG	GCAGCAGAAG	AAGATCTTTT	CAAATACACT	GAAGCAGGCG	30960
CTGATTTGGT	GATTTACAGC	GGCGCCAAAG	CAATTGAAGG	GCCAAGTGCC	GGTTTAGTGG	31020
TCGGGAAAAA	AGAATACATT	GACTGGGTGC	GCCTACAAGG	CAAAGGCATT	GGTCGAGCAA	31080
TGAAAATTGG	CAAAGACAAT	ATCTTGGGCT	TTACGCAAGC	AGTGAAGAA	TACTTGGCAC	31140

ATGGCAGTGA ATCAGGGGCT TCGATGCAAG AGCGCTTAA ACCGTTTGTT GAAGCCATCA	31200
ACAACTTATC AGATTTAACC GCCAAAATCG TCCAAGATGG CGCTGGCCGA GACATTTATC	31260
GCGCCAGTGT CAAAGTGGAC GGCAGAAAAA CCGCTAAGGA AGTCATCCAA GCATTAAAAG	31320
CAGAAAGTCC AGCCATTTAT ACGCGAGAAT ATCAAGCCAA CAATGGCATC ATTGAATTTG	31380
ACATTCGTTT TGTCAATCAA GAAGAAATGA ACAAATCGT GCAACGATTA CAAGAAATTA	31440
TGGACAAAAA GGAGAAATAA GTCATGTCAT TAAACCTAA TTATTAGAA GAACGCATTT	31500
GTTTAAACGT CTTAGCCAAT TCAGTAGAAA ATGCCCAAGC CTGTTATGAA GCAGCAGAAG	31560
GGCATGTTGT GTTAGGCGTC CTTTCAAAAA ACTATGAAAC CGATGAAGCA GCTATTGACG	31620
ACATGAAAAA ATACCAAGCA GCAACCAACA ACGCTTTATC TGTCGGCTTA GGGGCTGGCG	31680
ATCCTAATCA AAGTCAAATG GTGGCTCGCT TATCAGAAGT TCTACAACCG CAACACGTGA	31740
ACCAAGTCTT CACAGGTGTC GGCGCtTCGC GGGCCTTACT TCGTCAAGAC GAAACAGTGA	31800
TTAATGGCTT GGTTCACCC ACAGGCAAAG TTGGCTATGT CAACATTGCG ACAGGTCCGT	31860
TAAGCTCTGG TGCGCCAGCG GCGGAAGTGC CTATCGAAAC CGCCATCAAA TTGTTGAAAG	31920
ACATGGGGGG CAGTTCATT AAATATTTCC CGATGAAAGG CTTGGCTCAC AAAGAAGAAT	31980
ACCAAGCAGT AGCAGCAGCT TGTGCCAAGT ATGACTTTTA TTTAGAGCCA ACAGGAGGCA	32040
TTGATTTAGA AAACTTTGAA GAAATTGTTT AGATTGCGGT GGATGCCGGC GTTAAAAAAA	32100
TCATTCTCA TGTCTACAGC TCAATTATCG ATCAAGAAAC AGGGGACACT AGAACAGAAG	32160
ACGTGAAAAA GTTACTAACC ATGATGAAAA ACACATTGAA TAAGTAGGTG GCAAAGATGA	32220
AAATTGCTGC GTTTGGTGAA GTCATGCTtC GTTTCACTCC GCCAGAGTAT CTAATGTTAG	32280
AACAAACAGA GCAGTTAAGA ATGAACTTTG TAGGGACTGG CGTAAATCTC TTAGCAAATT	32340
TGGCTCATTT TCAACTAGAA ACAGCATTGA TTACGAAATT ACCTGCCAAT CGTCTTGGTG	32400
AAGCTGGCAA AGCAGCGCTA AGGAAGTTAG GAATTCAGA TCAATGGGTA GGAGAAAAGG	32460
GCGATCATAT TGGTAGTTTT TTTGCAGAAA TGGGTTATGG AATCCGCCCC ACACAAGTAA	32520
CTTATCAAAA TCGCCATCAG AGTGCTTTTG GCATCAGCGA GGCGAAGGAC TACGATTTTG	32580
AGGCCTTTTT AGCGGAAGTA GATATGGTGC ATATTTGTGG TATTCGTTA AGTTTAACTG	32640
AAAAAATCG TGATGCTGCT TTGATACTTG CTCAAAGGC GCATGCCTAT CAAAAGAAAG	32700
TTTGTTTTGA CTTTAATTAT CGACCTAGTT TGAATACAGC AAATAGCGCA CTCTTTATGC	32760
GTCAACAG	32768

## (2) INFORMATION FOR SEQ ID NO: 124:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6027 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

TTTATTTACA CCTGCATTGG AAGTCGAAGA TGCCGAAAAG AGTAGTTGGA CTTATCCCGT	60
TTATGGTTAT TTAGATAGCG AAAATCCTTG GGAAAAAATT GCGGCTTTAT TAAATAAACG	120
AACACAAGGC ACACCACGCT TTGCTTTAGA AAAACAAGCT CTTTCTTTAG CTCGTTTCGA	180
TCAATTGAAA ACCTATTTCC CAGCTAGCGA TTTTTCGCAT GATGTTACGC CTTTAATTGA	240
AAAACTTCAA TTAATTAAAA CAGAACCTGA AATTCAACGT TTACTAGAAG CTGGTAGTTG	300
GGCTGATGTC GCTTTTGAGA TTGGCTTCAA AGCAATTCAA GCAGGCGTTG CGGAACAAGA	360
AATTGTTGCA GAAATCGAGT ATCAATTAAA ACGTCAAGGT ATTCGCTCAA TGAGTTTGA	420
TACGTTAGTC TTAACAGGAA AAAACGGAGC TAGTCCTCAC GGGGTACCTG GCGAAACAAA	480
AATTGAGCCC CATGACTTAG TGTTATTTGA TTTAGGTGTC GTTCATAATG GCTATTGTAG	540
TGATGCCACA CGAACAGTCA GCTACTTGGA ACCTTCTGAT TTCCAAAAAG AAATTTATGG	600
AATTGTCTTA GAAgCACAAT TAGCAGCAAC TGAAGCGGTG AAACCTGGTG TTACAGCAGG	660
cGAATTAGAT GATATTGCCC GAGGTGTAAT TACmAAAGCT GGTATGGTG AATACTTTAA	720
CCATCGCTTA GGCCATGGCA TTGGAACCAC TGTCATGAA TACCCATCGC TAGTTCATGG	780
CAATGATTTA GTCATTGAAG AAGGTATGTG TTTCTCTATT GAGCCCGGTA TTTACATTCC	840
TGGAAAAGTA GCGTTCGTA TTGAAGATTG TTTACACGTA ACGAAAACAG GCTCCGAACC	900
ATTCACTAAG ACAACCAAAG AATTGCAAAT TATTCAATAG CATAAAAATA AAATGAGGCC	960
ATCAGTCAAC CGGACTGATG GCCTCATTTT TAAATTTTAA GGTAAGCCGC TGAACAAAAG	1020
CAACACGATC CGTTGCTAAC TTTCATAAGC TAAACGCTTG TTACCCTTTA AATTCTTTTT	1080
TTACGTCTTT CGTGTCATCT TTAACATCTT CTTTTGCTTC TGCAGCTTTT TCAGGAACAT	1140
CTTTGGTGGT TTCTkTTGCT TCATCGACAC CAGCAGAAAC AGTTTCTTTG GCCGCTGCCG	1200
CAGAATCTTT AACGTCAATA TAGATATCTT CTGCTGCATC TTCGGCAATT TCACCTAAAT	1260
CTTCTGCGTG ATCTTGAGCA TCGTCAGCTG CTTTTTTAAA TTGATCAGAT AAATCACCTG	1320
TTTGTTTTTT AAAGTTATCA AGCATGTCGC CAGAAACACC TTGTGCTTTA TCCAATGAAT	1380
CTTTTGTTTT TTCTTTGACA GAACCTGCCA AATCAGAGGC TTGATCTGAT AAAACGCCGG	1440
CTTTTTGTTT TGCGATTCTT GTTAATTCTG TTCCTTTTGG AACAGCATAA TCTGTGTAAT	1500
CTTGCGCTTT GTTTTTTAAA TCATCTGTTT GATTGATAA ATCATCACGT AATTCTTTAC	1560
CTGATTTTGG TGCAAGTAAT AATGCGGCAA CGGCTGCTGC TGTTCCACCA ATTACTGCyC	1620
CTAAGAAAAA TCCGCCTTTT TTAGCCATAA AAATCTCCyC CAATTAGTTA TtkGTTTTTT	1680
TATTTTTTGA ACGCATTGCT CTAAAAGCAG AACCGCCTAC TTTACTCACT nAAACCAGcT	1740
TTGCGTGTGC CTTTyCCTAC GGTTCTACC TTTGTAATCA AATGTTTACT TGAGGTGTTC	1800
AATTCAGAGA CACTTTCGCT TAAATCAGCA ACGGCCGTAA ACAACGGATC AATCGTTGCC	1860
ACTTTTTGAT TCACATCTGT CAATAGTTCA TTACTTTTAA CAAGCAAGCC TTCCACTTGG	1920

CGACTCAATA	CATTACATC	GCTAGTCACG	ACTTTGATTG	TTGTATTTGT	TTCATCAATT	1980
GTATCTTCAA	TTTTTGAAAC	AACTTGAGAG	ACTTTATATA	ACACAAGTAC	TAAAAAGATG	2040
ACTAAAACAA	CAAATGcTAC	CGCTGCAATT	AAGGCAGCAA	TTCCCCACC	TGACATATCA	2100
TTACTCCCCT	CTGAAATTTT	ATTTATAATC	tTAGGATAAC	ACCAACTTAT	ATTTTTCACA	2160
ACTCATAAAG	CTAAGAATGA	CAATTTATTT	TATAAAATTC	TCATGCCCTT	TTCTTTAAGA	2220
AAAAAAGCGT	AAAATTTCTT	TCTAACGAAT	TGAAATTTTA	CGCtTATTTT	TATTTATTTT	2280
TCTTTTGTTT	GCTCTTTTTT	CGCTAATTGA	GCACCTAAAT	TAATAATGTA	TTCTCGTAAA	2340
TTATCTTTTA	CTTCTGGATG	AGTCAAACCA	TATTCAATGC	TTGTTTTCAT	AAACCCAAAT	2400
TTATCGCCAA	CGTCGTAACG	TTCTCCTTTA	AACTCACGAG	CAAAAACGCG	TTGTGTTTTA	2460
TTTAAAGTGT	CAATCGCATC	TGTTAACTGA	ATTCGTAC	CAGCGCCTGG	TTTTTGATTT	2520
TCAAGTACAT	CAAAAATTTT	AGGGGTTAAT	AAATAACGGC	CGATAATCGC	TAAATCACTT	2580
GGTGCTTCCT	CGGGTTTTGG	TTTTTCAACA	AAATTCTCGA	CATTGTATAG	TCCTTTTTTCG	2640
ATTTCTTTTT	CTGGGTTGAT	AATACCATAT	TTTGAAGTAT	CTTCATGGGG	TACTTTCATT	2700
ACCGCAATTG	TCGAAGCATG	TGTTTGTTTG	TAATCATCCA	TTAATTGTTT	CGTTAATGGC	2760
ACTTTATCTT	CCATTAAATC	GTCGCCTAAC	ATCACGACAA	AAGGTTTCATT	GCCAACAAAT	2820
GCTTTCGCTT	GTAAAACAGC	GTGTCCTAAG	CCTTTCGGAT	GGGATTGACG	AATAAAATGT	2880
AAATTAACAT	CCGTGTGTTT	TCGACTAAT	TTCAATAAAT	CGGTTTTATT	TTTTTCTTTT	2940
AAGTTATTTT	CTAATTCAAC	ATTTGAATCA	AAATGATCTT	CAATCGGACG	TTTGGCTTTG	3000
CCTGTCACAA	TTAAAATATC	TTCGATTCCCT	GAAGCTAGTG	CTTCTTCTAC	AATAAATTGA	3060
ATCGTTGGTT	TATCAACAAT	CGGTAACATT	TCTTTTGCCA	TCGCTTTGGT	AGCCGGCAAG	3120
AAACGTGTTC	CTAACCAGC	TGCTGGAATA	ACTGCCTTTT	TAACTTTCAT	CTTTTTTCCT	3180
CCTAAAAATC	AAGTGTAAT	TCGTTTTCGG	TTTTGCCATC	GCGCAACATG	ATTTCTTTTG	3240
CTGCTTGTTG	TACATCTTCG	TCATTGTACA	AGACATTGTA	GATTGTTTCT	GTGATTGGCA	3300
TTTCGACTTC	TAGTTGTTGT	GCCAATTCAT	AGGCTGCTTT	AGTTGTTGAA	ACACCTTCCA	3360
CAATCATGCC	CATGTTTTCT	AAGACTTCAT	CTAACTTATG	GCCTTTTCCT	AATAAATTAC	3420
CCGCACGCCA	ATTTGAGAA	TGGACACTTG	TACAAGTAAC	AATTAAATCC	CCAACACCAC	3480
TTAAACCAAT	AAATGTTAAT	GGGTTTGCTC	CCATCGCAAC	CCCTAAACGA	CTAATTTCTG	3540
CTAAACCACG	AGTCATAATT	GCCGCTTTGG	CATTATCACC	AAAGCCTAAG	CCATGAATCG	3600
CTCCGGCGCC	CAAAGCAATA	ATATTTTTTA	AGGCCGCCCC	TGTTTCTACA	CCAATCACAT	3660
CATCATTCGT	ATAAATTCTA	AAGTAATCAT	TCATAAACAA	CTCTTGACAA	TACACAGCGT	3720
CTGCTAAATT	TTGCTTGCT	GCTGTAATGG	TTGTAATGTC	ATGAACAGCG	ACTTCTTCTG	3780
CATGGCTTGG	ACCAGATAAA	ACCACAATCC	CTTGCCGTTT	TTCTGCTGGA	ATCTCTTCCG	3840
CTATCACTTC	TGAAATCCGC	TTATGTGTAC	CTTGTTTCGAG	CCCTTTACTT	GCATGAATAA	3900

TAATTGGTTT	TGTTTTTAAA	TGTTGGGCCA	CTTCTTGTGC	TACTGAACGA	ATGGCTTTAG	3960	
TTGGTACCAC	AAACAGTACC	GCATCAACAT	CTACTAAAGC	AGTGGCTAAA	TCAGTGGTTG	4020	
CCTGAATTGA	AGTAGGTAAG	ATTAATTCTG	GTAAATAATG	TTTGTGGTG	TGTTTGTATT	4080	
TATTTTCATCG	ATTTGCTCTG	GTTTATTCCC	CAAATACAGA	CCTCGTGTCC	ATTTcGGCTA	4140	
ATACTgAGCT	AAAGCTGTcC	CCCAGGAACC	TGGACCTAAA	cAGCGACTTT	TgTTTCATCG	4200	
GTGAAACTCT	CCTTTCGCTA	AACAACGTTA	CTTCTTTAGT	TTACCATATT	CTTCAATCAA	4260	
GTTGGATTAA	GAAGTTGTTA	TTTTTTTCTT	TTGTTTCTCC	CGATTGTAAA	ATTTCAAATT	4320	
GGGATTTCCT	CTGCGAA	GA	TCACAATAAC	AATTGCTGCC	ACGAACATCA	CTAAAGACAA	4380
CAATTGTGAG	ACACGAATGT	TACTAAACGC	ATACAAACTG	TCCATCCGTA	AGCCTTCGAT	4440	
GAAGAAACGG	CCGAAAGAAT	ACCATATTAT	ATAACCTAAA	AAGACTTCGC	CTTCTTTTAA	4500	
AAAGTGTGGC	TTTTTGCGTA	ATAACACTAA	AACGATAAAG	CCTAAAACAT	TCCAGACAGA	4560	
TTCATATAAA	AATGTTGGTT	GATGATATGT	TCCGTTGATA	TTCATGTTAT	CAATGATAAA	4620	
CGTTGGTAAA	TGGAGATTTT	CTAAAAATTG	TCTGGTGGTC	GCTGGCCCAT	AGGCTTCATG	4680	
ATTCATAAAG	TTGCCCCAGC	GACCAATTGC	TTGCGCTAAA	ATCACACTTG	GTGCGGCAAT	4740	
ATCTAAAAAA	GTCCATGTAG	AAATGAAACG	ATGACGCGTA	AAGAAAAACA	GCGCCaGTCC	4800	
CCCACCAATA	AGACCACCAT	AAATCGCTAA	GCCACCATT	CGAGTAAAGA	AAATTTCAAT	4860	
CGGATTGTCT	ACATAATCTT	GCCATTGAAA	AGCGACGTAA	TACAAGCGCG	CGCCAACAAT	4920	
TGCTGCAGGT	AATCCCCAGA	GCATAAAGTC	AAAGACATCA	TCTTCTTTTA	AACCAACACG	4980	
AACGGCTTCA	CGACTACTTA	GCCAAACAGC	TAAAGCAATC	CCTGAAACAA	TAATAATTGC	5040	
ATACCAATAG	ACGGAATAC	CGAAAAGGCG	AAATGCAATT	GAATTTACTT	GGGCTAACAT	5100	
CATCACATCT	TCTCCTTCTA	TTCTCCTGAA	TTTTCTTCAA	TCAAACGGGT	TAAGCGCTCT	5160	
TCGAACGTTT	TTGTTGCATC	ATAGCCCATT	GTTTTCGCGC	GGAAGTTCAT	TGCTGCTACT	5220	
TCAATGATAA	TGGCCACATT	TCGGCCGGTC	TTAACAGGAA	TACGAATCTG	AGGGACATCA	5280	
ACGTTAGCTA	TTTCAACCAT	TGCATCATCA	CTACCAAGAC	GATCATATTT	CTTGTCTTTT	5340	
TCCCATGCCT	CTAAATAAAC	GACCAATTGA	ACTTGCTATA	AGCCACGGAC	CGCACTAGCT	5400	
CCAAACAGGT	TCATGACGTC	AATAATTCCG	ATACCACGAA	TTTCAATTAA	ATGCTGCAAA	5460	
ATTTTTGGTG	GTTCTCCAAC	CACGGTTAAT	TCATCTTGCT	GATAAACATC	GACGCGATCG	5520	
TCTGCGATTA	GCCGATGTCC	ACGTTTAATA	AGCTCTAAAG	CTGTTTCACT	TTTACCAATA	5580	
CCGCTATCTC	CTTGAATCAA	AACACCAAGT	CCATAAACAT	CAACTAAAAC	TCCGTGGACA	5640	
CTTGTCCGAA	CAGCTAAACG	GCCATCTAAA	TAACTGGATA	GTTCCCTAG	TAAACGGGAA	5700	
GTTGAAATCG	GTGAACGTAA	TACAGAAACG	CCATTTTCTT	TTGCTGCTGT	AATTAATTCT	5760	
TCGGGGATTT	CTAGTCCTCT	GGATACTATA	AAGGCTGGAG	TATCTTTCGC	ACATAAACGG	5820	
CGCATCACTA	ATAAACGCTC	TTCAGGCATC	ATTCGTTCCG	CAAAGGTAAT	TTCTTtGCTC	5880	

CCAAATAATT GTAAACGGTC ATGGGAATAA TAATtArAAT ACcCaGTAA TTCTAAACCT	5940
GGCCGGGAAA TTTCACCAGT TTTAATTGTT CTATTTAAGC TtCTtCATC TCCGTATACT	6000
ACTTCTAAAG AnAGATTTTC TACCAGT	6027

## (2) INFORMATION FOR SEQ ID NO: 125:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4915 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

CAACAGTCCC AGATGATAAG GTTCTGCTAC CATTACGAAA CTCTACATCA GTATATAGTT	60
TATCTGACAC TCGCATAATA ATTTGGTGCG GATTTTCAGT AAAGTAATTA TCTGAATTTG	120
ATACCAATTT TGTAACATAT GATTTGTTAG CAGTTATATA ACCTTTGGCA GkCyTTAATC	180
yAGGTAATCC GTTTTCGGTA TAATCAACAC TaAAACTTgA TTAATGTATC TTTAGAAACA	240
AACTCACtTT TTTACTGAAA TTTACATCTT GATAATAATA ATCATTGTT TTCATAACTA	300
CATACTCTGG ATAATCATAT AGATAATCTT TGATGTTTG CACAACTTT TGTGCATAAC	360
TTTTATTAGC CGTTAAATAA CCATTTTGGG TAATCAAACG AGGAACTGA TTTTCACGGT	420
ACTCGATTCC CACAACAGGA ACTAACTTC CCTTGGAAC AGCTGCTCCT TTATATAAGA	480
AGTCTGAATC CCCATAAAAA TTATCACTAT TTTTCATAAT AATTGCCCCT GGATTAACCG	540
TATGATAGTT AGCAATCCAA GATCCAGTAG CTACTACATA GTTCTTGTTA GCCGTGAAAT	600
AATACCCATT TGCCATCTTC AATCTAGGAA TTCCTGATT TGTTTTCTCA ATAGCAGTTA	660
CTTTTAAAC ACTGCCTGCA GATACTTTAT TTGCTTTTGT CTTGAATTCT ACATCTTTGT	720
AATAGGTATC TTCATTTCTA GTAATTACTT GTCCAGGATT AGATGTGTAG TAAGTATCTT	780
CTAATCCTTT ATATTGTTCA ACTATATTTT TATTTGCAGT CAAATATCCT TTCGCTGTTT	840
TCAATCGAAA TATACCGCTT TGAGTATAAG CTAAATCTTC AACTTCTACT ATTGTTCCCT	900
TCGAAACCTT AGACAACCTT TGCTTAAATT CTGGATCAGC ATAGAAATAA TCATCGCTTT	960
TCAAACGAAC TTTTTTAGGA TTTGTCGTAA AATACAAATC AATATTGCTT TGTGCCGCTA	1020
ATACATAATC TTTATTTGCT GTTAAATAAC CTTGGTCTGT GACCAATCGA GGAATTCCTG	1080
CTGAACTATA TTCAATCCCC TTAATGGTAA CTAGTGTATT TTTTAACT CGCCAACCAG	1140
GAGTTCTAAA CGCTACATCT TGGTAAAAAG TATCGTCATT TTTTCATGATC ACTCGACCTG	1200
GaTTAGTTGT GTAATACTTA CTTAAGTcAG GCCCACTACT ATTACCCATA TTAAaAAAT	1260
TACTATAGGT CwGGAAtATC AAACGGACGC CCTTcATAAT tAGCTAAACC AGGgAAATAC	1320
ATTTGCGAGA CCATTGCCAA GCAccTGgTC ATTATyCCmC TGCATGGAAC TATCTGGaGT	1380
ATATGGATAT tGCGCTACCC aTACTCTGTC TCTTCCgAAG GCGGAGGTGT CGATATAGCC	1440
ATTkGTTAac CACCATTTCG CAACATACAA CGCATCATTT TTATACCCTA ATGCCTTTAA	1500

TTGCTGGTTA AACGCAACAG AATTAGcATG cGTTTCGCCC ATTGTTTGTA AGCGTTGGAT	1560
CTTCCGCATC ATTAAACATT ATGGTATTTT TATCTAAGCC AGAGTTAGCC GCCGCCTGAG	1620
CAAAATATCT TGCTTCATCT TGTGtGTCGC AGCAGAGACA TACATACTAT AGTGATATGC	1680
TGAAACTTTT AAGCCCGCAG CTTTGGCATT TCTAATTTGG CCCGCAGCAT ATGGATTAAC	1740
ATACCAAGTT CCTTCTGTCA ATTTAACCGA AACACCTGTT ACACCATAGC TTTTAATTTT	1800
TTGATATTCA GCTACTGAAA TGTCTCCATT CCAACTAGAA ACGTCAATAA AATTCTTGCC	1860
AGGCAAATTA GGATCATTCG CATAGACGTT CGCAACTGTT GCTCGCGCTA CCCTAGAAGA	1920
AAGGCCGTTT TTCATAGCAT ATGTTCCTTT TTTATGGTTC GTTCCTTCAC GTGCTGCTTC	1980
TTGATAATCC GCTTCATCTT GGTCAATGGT CTTTTTTTCT ACTTCAGAGC TTTGTGATGC	2040
TTCTGTTTTA CTATCTGTTG TGTTTGTTTC TGAACAGAA GAATGTTCTT CCGTTGTACT	2100
ATTTGTTGAA GAATCAAGAG ACTCTTCTTT TGTGCTTGAA GTAGAACTAT CTCTTTCAGA	2160
GGTATTTTCT TCTGTAACAG AAGATGTATC CTCCGTTGAA TTAGTCGTTG TTTCTGACGT	2220
AGCTTGAGTG ACCAAGCCAG TcTCCCCCGT TGTTTCTGTT TTTACAAGCT CTTCTCCTGA	2280
TGCAATCACA CTAAAACTAT TTAATAATAC AGTACAAATT AACATACCTG AAATGATTTT	2340
TTCATGAAAC TCTCCTTATA TATTTAATAA TATGTATCTT TCTTAAACAC TTGCCATTGC	2400
CTCTCGAAAA ATGTGTATCC TATAACCTTC AACTAATTGT TTTCACATTT TAAAACTAAC	2460
ACCAATTATT ATCTCATACC TTTATTTCAA ATTCAATTAC GCTTTTTTAA CATTCTTATA	2520
TCAATTTTTT ACTTTTAAAT GGTAAAAAGA TGA AACAGAGA TCTTTAATCG ATTGTTTTCA	2580
TCTTTTATTT ATTTTTTAAG ATATTCTGTG TACGATTTTT TAGTATCAAC AATTAGCGCT	2640
TCATGATGGG GTTGGCGATC TTCCTCTGGT GGCAATTGCA TATAATCCCC AAAAACCTCT	2700
GTTAAGTAGC GCTTGTAGCC CACTGGTATT GGCATTTTTG TATCTTCAAA TGGAACAAAA	2760
ATAGCAGATT TAAAACTTC CAAATGATAA ATATTACCCA TGTAGCGTGG TCCGACACAT	2820
AATTCTGTAA CATAAGGAGT GGAGCCGAAG GAATACTTAG TCATTCTTTT TTCAGCAAAG	2880
CGCCAAATAT GATAGCGAAC CTTTTTCGAT GGAAAGATAT TCAATAATAC CCGACTCCCC	2940
GTTGCTAAAA TGCCACCATG TTTTCTGGT ACCACTGTG AACAGAATAA AGCATAAATT	3000
AATGCCCAA TTTTTGTTT CTTCCTTTGA AAATTTCCTT TTGGCGCACC ATCTAAAGGA	3060
AAAATATCGA TAGGAATACC ATGAGGAATC GCTAATCCAT CTTGATAGGT TTTAATAAAC	3120
GTTGTCTGTG CATCCCTAAT TGTCAATAA GAATTATGAT CATTATAAGT TTTGAAGCA	3180
ACTAACAATG GATAACATTC TGTATCTGCA TATTTAGGCC AAAGCACTTT CaATTTTTCa	3240
TAATCTGGTC TTGGCATAAA GAAATCTAAG TCATCATCCC aTGGAATAAA ACCTTGACTG	3300
CGGATTGCTC CGATACAACC GCCTCCACAA AAATAACAGG TTAGATTATG TTCTTCACAA	3360
AAAGCCACAA AATATTTTGA CATGCGTAAA ACGACTTGTT GAATTTCTAC TAATTGCGTC	3420
TGTGTTGTTT GTTCCACTTT TCCTTTTCCT CCCTATTATC TCTTGCATA TTAAATGTT	3480



GTATAGAATA AAAAACGCGAT TAAGTAGAAT GAACCTTTTAA TAACTGAAAA CTGTTTCATTT	3540
TGACGCAAAA TTTTCCAAAC CCAACTGATT GTATTCCATT TGTTAGAAGA TGTAGAGGTG	3600
GTTAGCTTAC GATAAATGGC TAATTTTTTCG TTAATAAAAT AGACCGTATT TATTTCTTTC	3660
AAAATGTTTA ACCAAGTAGC ATAGTCTTCA TGTTTTAAAT CCGGCATTAA ATGGCGTTTA	3720
AGCAGATTAG ATTCGACTAC GACTGTTAAG CAGCCGATAG TATTTGTTTT TAAAAGATCT	3780
TGATAGGTTT GTGGTCCCTC ACTAAAATTA CGCTCATGAG TAATGTTGCC ATTTTCATTA	3840
CATACAAAAT AAGAACTACA TGTAATAAA AGTTTTTTTT CTTCCATCCA CTTGACTTGA	3900
ACTTCTAATT TATTAGGAAG CCATAAATCA TCACTATCTA AAAAGGCAAT ATATTGTCCC	3960
TTTGAAGCTT CAATTCCTTT ATTTCTGGTC TTTGCAATGC CCATGTTCTG TTCATTTTTC	4020
AGTAAAAAGA TTCTTGAATC TTTTTCATA AAGGCAAGyG CAATATCAAT AGAATCATCT	4080
TTACTTCCAT CATCAATGAG TAATAATTCC CAATTCTGAT AGGTTTGTTT CAGAACGCTT	4140
TCAATAGATT TTGATAAAAA TTTGCCAGCA TTGTACATGG GCATAATAAT TGAAACAAGT	4200
GCTGTTTCCA TktGTTCACT CCTCGTTGCA TTTTATGTAA GAATGTAACA GTTTATATTA	4260
TCGCATAGCT CTTTATTAT TTCAACGTTT CTTGTTTCCA AAAAAAGTTT GAGACATAAG	4320
CTTAAATAAC TTATGTCTCA AACTTTTAAC GGTAAATTCT AATAGCATTC TCTTTACAAT	4380
AATCCATCGT CAATTCTTCA CGTTCCTCAG ACTCATCCAA TCCTCTtGAT GATACTTTAT	4440
CAAAGAGAAT TTTGATGGTT tgAAGCAGAA TTTTCATATC CaATAGCAwG aATAATTCTT	4500
GATGTACAGT AAATCAAAGT TCAATTTACT ATTGTAGTCA GAAGCATACT TCCCATAAAC	4560
TTGGGCATAG CCAGTAATCC CCGCACGTAC ATTGTGACGT AAATAGTAAT TCGGGTTTTT	4620
CTGATTAAAT TGATCAACAA AGAATGGTCG TTCAGGACGT GGTCCGACAA TGGACATATC	4680
CCCGCGAATA ACATTGATTA GTTGTGGCAA CTCATCAATA CGTAAAGCGC GCAAGTATTT	4740
ACCAACTGTA GTGACGCGAC TGTCGTTACT TGTAATAAG ACTGGGCCTG ATTTTGCTTC	4800
TGCTGTTGCC GACATCGTTC TAAATTTTAA AATCGAGAAC TCACGTTGAT CTAATGTAAT	4860
CCGAGTTTGT TTGTAAATCA CTGGTCCCGG TGAGGTTGCT TTGACTAAAA TGCGC	4915

## (2) INFORMATION FOR SEQ ID NO: 126:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3149 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

ATTAATGCAA TAGCTGCCCC TACAAGCGCT ACTCCTAATA AAGGTATTTT CATATAAACA	60
GCCATTACAA AACCAATAAC AAGATAACTA AAATATTCTT TTGCAGGTAA ATATCGCAGT	120
AGTAACCCTA TACCGACAGC TGGTAGTAAA CCTCCAGCAA CAGTTAATCC ATTGGTTAGC	180

CATTCTGGAG TATACGATAG AATTGTATTA ATTAATGATG GA <sub>CCG</sub> AAAAT TACTGCCAAT	240
GCTACTGGTA TGCCTGTTGT TGCTGCTGTT AAAAACACGC CAAACATTTG CATT <sub>TT</sub> GGTA	300
ATCATACTAT AATCACCTTT TTTCACACCT GCTTCAGCTT TGTGAACCAT CCAGATACCT	360
ATGGTATTTC TGATAACATC TAGTTGGACC ATCAATAAAG C <sub>T</sub> ACTGGTAT ACCAATAGAG	420
ATTCCTACAG ATGGTTTTTG CCCAGTAGTA ATTGTAATAA AAGTTGCTAC GATACTTGCT	480
GTTTGATAGT CAGGTATTGA AGCTCCTCC <sub>c</sub> AAAATTACTT ATTCCCaAAG ATAGTAACTG	540
TAAAGTACC <sub>a</sub> CCGATAAATA AACCTGtTGA TATATCmCCt AGGATTAATC CTGTAACAAA	600
ACCCGCTGnT TACCGGTTGA TAAGTTCCAA ACaTTGTTGA GTTTTTTCA AAGTTAATAA	660
AAAATCCATA CAGCACAATT AATACTAATT GAAAAACT T <sub>GCT</sub> TCCATT AATTTTCCTC	720
CTTAATTACA TCAATAAATT TTGAAAATTT ACCTTTTTTCA CTGT <sub>TTT</sub> TGA TAAAAATCGG	780
GATTTGTTGA ATGTCTGCTT CGACTACCAC TGTTTTTCCA CCATCAAAGT TTTCTCTGT	840
CCATAGGTTT GTCCAAC <sub>TTT</sub> CACCTTTTGG TAAATAAACT TCTCGACTAG TCTCTTTATA	900
ATTTTCAATA GGCGCAACAA GAATCGCATC TCCAAATAAA TAGGTATCTT CAATATTCCA	960
AGCATTCTGA TCATGTGGAT ATTCATAAAA TAATGTTCTC ATAAGTGGAT CACC <sub>AA</sub> ACTC	1020
ATGGGCCTGT CTCATCAACT TACTAATATA AGGCTTCAAT TCTTCTCGTA tTTTTATATA	1080
TTTCGtTAGT ATTTGTTCGA CTTCTTTACC ATAAGACCAA ATTTCATTG GAGCTCCGGT	1140
AACCATTGAT CCTCCTCCAT GATCAGATAA TGGTTTGCTG TGTGGTAAGC GATCACCATG	1200
CATTTCGTAA ATAGGCGAAA ATGTTGCATA TTGGAACCAG CGGACTAGTA ATTCTCTAAA	1260
CTCAGGATCT TTAGGATCTC CTCCATGAAA TCCTCCAATA TCTGTGTCC ACCAAGGTAA	1320
CCCAGCCAAC CCAACGTTTA ATCCAGTATT TACTTGATTA CGAAAAGCTT CAAAGCTAGA	1380
ATCAATATCT CCTGACCATA ACAAAGCTCC ATAGCGTTGT GCCCCTGCCC ATCCACCTCT	1440
TACAAGTGTA ACAACTGATT GATTATTTCC TAAACCTTCA GAAATCATTG ATAAATAATC	1500
GATTGGATAC AAATTC <sub>CAA</sub> TTTGTAAATC gGTTCC <sub>TTTT</sub> TTATATCGAT AATTATCGAA	1560
ATCATATACA GAATACCCAG GCTCAGCTAC ATCTAGCCAA AAATATTTAA CTCCTTTGTT	1620
AACATAATTC TTTTAAATGA GGTTC <sub>AAAC</sub> ATATTCTCGT GCTTTCTCAT TAGTCATATC	1680
AACAAAAATT GTATTTCTT GAATTTGCAT TGTTAATCTA ACGCCACGAT TTACTTTCAC	1740
TAAGTAACCA TTTCTAGAT ATTCATTAAA ATTCATTGCA TCAGTTTGTA CAGTAGGCCA	1800
TATGGAAATA ATGGCTCTA CATTATATTC ATCTCTTAAT GTcTTAAAAA TTGTT <sub>CAG</sub> GT	1860
TGAGGCCAAT ATnCTcATCA AATCGATATT CACCTTGTTT AGGCCAATGG AAATAGTCAA	1920
TAGCTATAGT TGAAAGTTTT ATTCCCAAAG CATAGTATTT TTTTACAACA TCCATTACTT	1980
CATTTGGAGA TCGATAGCGT AGTTTACTTT GCCATAATCC CAATAAGTTT TCAGGCATT	2040
TAGGAACTTT ACCAGTTAAA TCACTATAAT TTCGaAGTAT CTCTTTTGGG GTATCACCTG	2100
CAATTACTAC ATAGTCAATA TAATTAGTAG AAGCCATTTT CCATCGAGTC ATGTTTTTAG	2160

CAAATGATAC CTCTCCTATA CCTGGATTAT TCCATAAAAA TCATATCCT AACTCGATA	2220
CATAGAATGG TATAGATACT TGTGAGTTTC TTTGAGCTAA TTCTAGAACT GTATTCTTCA	2280
AATCAAGATA AGATTGTTGA TATTGTCCCA TTCCAAAAAT CTTCTCACTT GATTCAGATT	2340
CAAATCTAGT TGTAACCTGA AAACCACCTT GTAAATTAGA GTGGTATTCT CTAGACTTCA	2400
ACTTCAGCGT TGAATTAAAA TCTTTAGTTA TTTCAATTGT ACCTGCATCT TCACTTCCGT	2460
CATCATGTTT GACAGCTCTC AATCTGATAA ACTCTTTTAA TAGCACTTCA TTTTTTTCAT	2520
TATAAAAAGT CAATCGGTCC CGATGGTCTA ACACAACTCT AATTTTCCCG TTTATTAATT	2580
CTGTCTTATT TTCGTCTTGA TTTATTATGA CTGTACCTAA AGCATCAGGC TCACCAATTA	2640
AAGAATTAAG TCTATTTTCA AAGTTCATAT CTACAAACGA TCTGACTCTA ACGGCATTTT	2700
CACCCCAAGC TTCTACAATC AATAATTCAT CATCAAAAGT TCTCcGAAAA AGATTTTGGT	2760
GTAACCTCAA TCACATCTTC CAACCCACTT TCAATTTCTT CTTAATTAGT TTGGTATATG	2820
ATAACAATAC TAAGTAATCG CAACAATACA AACGCTTTCC ATACGCACGC CTTTGTCTCT	2880
CTCTGTGTcm ACAAAAGGTT ATAGAtTAAT TTGTTTTcmC GAtAAACTAA TTAATAAAA	2940
AAATAAACAT AAGATAATAA GCATCAATGA CTTGAAATTT ATTCGTTAGC ACGTTAAACT	3000
AATTAATAAA TAGACTATAA AACAGGCTCA GTAAATGTC AACTGTAAAC GGCTATATTT	3060
TTTTTGATA AATTAATTGG CTTTAGGCAA TAGGGAATCT TAACCTTAAT TnCACTAATT	3120
TATTCnAATT TAGGGGGGAA ATGGTTTCC	3149

## (2) INFORMATION FOR SEQ ID NO: 127:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7019 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

ATTcCCCkGC tTCyTTTyCC ATAAAAATAC yCTTTtCTAG ATATTTTAAa GCATTTTaTr	60
GAAACGTTtC ACCAAAATAT GAACATCTTT TTCTTAAAAA TAATTCTGTT AAATAACAGA	120
AAAAAGTCTG GTCAATCCGA CCAGACTTTT TAAATATAGA TAGTAAATG TGTAATTTCA	180
GGGATCACCC CAAAACGGGC ATGAATGGCG GTCGTTCTTA AGCCTGTATT AACATATAGT	240
TGCGTTTCTk GGTCCAACCTG ATACAATCCT CTGGTATACT TTTCTGCTAA AACATTTGTG	300
ATTGGATAAA AAGGAATCCA AATTTGTCCG CCATGGCTAT GTCCTGCTAA AACTAACTGT	360
GTACCCGTAC CAACAAATGC ATCCGCGACA TCTGGTTCAT GGGTCATTAA TAACGCATAG	420
TCATACGTTT CTCGATAGGA TAATGTTTCT GAAACGGAAG GATTTCCCAA TAACGAGTCA	480
TCTAAACCAC CAATAAAAAG ACGGCGTCCA TCTGCTAAGG TTAAGTTTTC TCCTTGTTT	540
TTCATAACGG TAAAGCCACT GGCCTCCATG aCTTCrnCAT ACACCGCAGA AGCACCGCCA	600
CCATAATCAT GATTTCCCCA AACTGCATAT TTTCCAATTG TCGCTTGTAa CTGTTTAAAC	660

TTTTCAATCA TCGGTTCTTT TTGTTGCGGA TTTTGGCAT AATTGTCAA TAAATCACCT	720
GTAAAAACGA TAATATCTGG TTTTTCACGA TTGACTTTTT TTATGACTTT ATCTAACCGC	780
TTCTGTCTCAT AAAATTCTGA AACTTGTATA TCCGAAAGTT GCACAACATT CAATGGTTCT	840
CGTTGATTTT TCTTTCCCAA AGTCACCTGg ATTAACGTGa ACTAAATTCG GCTCAATCTT	900
CCACGcATAG ATTGGGAGTG CGaTAATTAA GaAAAAAAT AAGAGCATTA TACTTAAACC	960
AATTTTTTTTA ACCACATCAC TCGCTCCTTT CTTTCTTTG ATTTCCGAAG ACTTAGTATA	1020
GAGAAAAATC TTCTTAACCA CAATCTAAAA CAAAGGAAAT TAAGGCTTTG TTAAACAACC	1080
TAAAGAATTT CCACAATCTC TTATCCTTTT CTTAGCTTTT TTATAAAAAA AGCTAGATAA	1140
TTTCATCTGT CGATGAAATT ATCTAGCTAG AACATTCTTC TCTTAATGAT GATGTTTTCG	1200
AAACGTTTGA TAACCAATTA AGCAAAATAA AATCCCCAAT AAAATCATTC CAGTTACTGA	1260
AAATAAGCCT GCTAACATCT TTCGCATCCT CTCTGTTGCT AATCTTCCTT ATTTATAGTC	1320
ACTATTTCTT GTTGTAAGT TGGTTGAAAT AGACGACGTT CTGCTAATTT TCTTGcTAAA	1380
ATTGCTTCAC TTTTTTTCGT AAAGATACCT AAGTTATGGC GCTTGCCTTT CAAAGAGATA	1440
TAAGCCAGCC ATTTCCCAGA AGCGGTCTGG CTCACCCCTT TGTGTCCACT GCTATTATTT	1500
TTGGGGCGTT TACTTTGTAA CTTCCTGGA ATTACGCCAG CAACTCTTTG TCGGTAATCC	1560
TTCTTTTTTAC GATGACCGCA ACTTTTGGTA TGGCCATTCA ACAAATTTTC ACAGCGAACC	1620
CAGCAAGATT GTCCGCACGA ACACAGACAA TACGCTTTCA TTGTCTGTTG AATACTTTCA	1680
ATCCGTACCA CGACTAATTC ATGATAGTTT TTTTCAACA TTGTTTGATA GCGGTGGTGA	1740
GCACATTTCTG TTCGATACCC ACGAATTAGT TGTCTTCTTG AAGCTTCTGC ATAGCTCCCA	1800
CAAGAACAGC GACAGGCCCA AACAAATCAGC CCTTCTTTTC TTTTGGCCGT TGGTGCTACT	1860
ACTGTCAACT CACCAAACCT TTGGCCAGTT AAATCTTTAT AGTTACCTGC TTGTCTCAAT	1920
TGACTATCCT TCGTATTCTT GTTCTCGCCA CATCTCGATT GGGTCTCTCT CATAGTGCTC	1980
ATTTTCTGA GTCTCTCCCT TTCTTAGTTC TCTGACCACC TCAGCAAATA CTTGCCAATC	2040
TTTTGTTTCC GGCAAATGAC GCCATGTCAA TTGACGACAC CGTACTTTCG AACTATAAAA	2100
ATCGGACAAA TACAAATCGG TTTCTAAGGT TAATTTGTGA TCAATCATGA CATCCATATA	2160
ATTGAAAGAA GCAATAATTT TTGCAATAAA GCGGTTATAA TCCTCCCCAC CAGAAAAATC	2220
AACAAATACA TGAATCGTTT TTTCAATTGC GCAAAAAGAC CGATCATTTA ACAAACAAAA	2280
CATAAAATCA TAGTATAGAT GCACGCGTTC ATACATTAAA AATCGTTTTT GATACAGCGA	2340
CTCTGTCTTT TGTATAAAAC CATCTAAGGC GCGATGAATT TGAGGAAACC GCTCTTCAAA	2400
GAAACTGCTC TGCTTGCGG GAATAAATGC AGCCACCGAA AAGTATAGGC GGTCCCAGCG	2460
ATAACAAATC AACTTTAATT TTTCTTTTGT TTCCACCGTA ATTCGCTTTT CAAAGAAAGG	2520
AATAACACTA TTCATATGAT CAATAAATTC GTCTAACATG GTTAACTTAT CTGGAAAAAC	2580
AATCGGACTA TCAATGTATT CATTTAAACT TAAAAAAGA AAAATATATG AAGTCTCTTT	2640

GCCAAAATCC	TCTTTAAAAA	GAGGGGTTC	GCTTGATAA	ATTTTTTCAA	ACGCATGCTG	2700
TAAACAGCTA	GGCACATGAA	CTCCGTAAA	AGAGTCCACT	ACATGATCAG	AAAATACACG	2760
AAAACGTTGG	ATCGTTAAAA	GAATAAATAG	TTGTGCGCTT	TGATTAAGAC	TAAACGTTCT	2820
TTGATACATC	AGTTGTAAAT	CGTTTATAAA	TCGTTTAATG	CCAGGACTAT	TTTCCAGCAA	2880
AATGTTAGGC	AGCCTTTCTT	GATTACTTCC	ATAAAAATAA	TACAAGCATT	GAAAACAAAA	2940
ACTTCGGATA	TTTTTCTCTT	CTCCAACCAA	GCCTGTTGAC	TTTAATTTAA	TATGCCAATT	3000
AGCTAAAATC	AGATTTAGTT	TATACGAGAA	GGCATAAATT	TTGCTTTTGC	TCATAAAATG	3060
AATTTGAGAA	AATTCTTGAA	TCGTGTGCTT	TTGATAGAGG	CCCACATCAA	GAAGCAACTT	3120
GGCCATTAGT	GATTGATCAA	AATAATATTT	TCTTAAGGCT	TGATAATCAA	TTTGCGATAA	3180
TTTTTGCGCA	CTAATCAGTC	CTTTAGCTGG	AAAAAGAATT	TGACTCCTTG	AAGAAATCTG	3240
TGTACATTCA	TAATTTATTT	GACGGCACAA	CGATTCTAAC	TTTTGAGTGG	ATAAATCAAA	3300
ATATCCATAA	AATTCTTCTT	TTAATAATGG	ACCTTCACTA	TTCATAAAAG	TTTGATAAAT	3360
AAGTAATTTA	TCTCGGAATT	CTTGGTCTAA	TAATTCATCT	AACAAAGTAG	TCACATTTTC	3420
AACTCCAAAC	TATCACGCCT	TAAGTATCTA	TTTTTTTAAC	AAGACGTGAG	TGTCAATCTT	3480
GCTTATTTTT	TCCACAAATT	TAAAAATGAT	TCTTTCAATA	TTAGGAATC	TTCATAGAAA	3540
AGTATTATTT	TAACCAAAAA	AAGCCCCCTT	GCGTCACCAA	GGAGGCTTTT	TTTGTTATTT	3600
TTTTACTTTA	AATTTAAAGA	AAAGCTGATT	TGTTTGACTA	ATTTTCTGTT	TCTTTTCTAG	3660
CCTATTCGTC	AAAAATGATT	GAAACTTTTG	GTTGCTCTTC	TATACTATCT	TCCTATAAAG	3720
TTCTTACTCT	TTTAAGTATC	AGCTTCTGTC	AATAAAAGTC	CTCATTTGGC	ATTGCTAATC	3780
TGAATACAGT	ATGACTACTT	GTTGCCGCTA	TTTTTTGCTG	TTATTTTACT	CTTTAATTGA	3840
AGCGCTAGAA	ACGCTGTCAA	CTGTTGGCTT	TTCCGGTCGCT	GACTCTTCTT	CATGGACTGA	3900
TGCACTTGAG	ACACTATCTG	CCGTTTCTGG	CGCTTCTTTT	TTGGAGGCAC	CACTGACTGC	3960
ATCGACGTTT	TTTCCACCGA	CTAATTCTTG	CCCAGTTTGT	TTTAAATACC	AATCAACAAT	4020
TGGTTTAAAG	TCTAAAATAT	ACTCATTAAAT	TCCCGCGTAA	TTTCCATTAT	CATCATACAT	4080
CGCTTGATAG	TTATGCACTA	CATAGGTATC	TGGTCCGTGC	ATTGGAACAT	GAACGCGAAT	4140
AGCATCTGTT	TTTCTGACC	GTAATTGTTG	AATCACCCAT	TCTACATTTT	TCAATGCTTT	4200
TGGCGGATGG	CAATTCGCTA	ATGGATTACC	AACTTGCTCT	GGTTGCCGTT	TAGCAAACAT	4260
TTCTTCAGCA	GCCATTTTTT	TATTGTAATA	TAAAAATTGA	TTGTTACTGT	CTGCATAAGT	4320
CAATTCCATT	GGCATTGATT	TTAAGAAATA	ATTTAGCTGA	TCTACGGTTA	AAATGCCTCG	4380
ATCTAATTTT	ACATATgTgt	CGCCTTCCAC	AGCCTGAAcC	GctTCTGCTG	CTTgtTCTAC	4440
CCAATCATCT	GCTGCCATAT	CAACCCCTTC	GATCGTTGTT	TCAATGGTTC	CTTTTGCATG	4500
TAAATCCTCA	GGTAAGTCTT	TTGGCTTAGC	TCCTTCTAAT	TTTGAAATCA	CATCAATGAA	4560
TTCCaCGAAT	TTTTGTAATG	TGCTTTCTAA	AAATTGACGT	GTTCCCTCCG	CAATTAAGTT	4620

GCCGTTTTCA TCAAATGCTT CTTTAGCTTT CCCTAATAAA AACTCGTTCC CaGGCATCaC	4680
GATTGCATTG ACACCAGGTG CATCAAGAAT TTGACGGAGA TGTAATTGCG CACGCGAGGT	4740
TCCTTGATCA TAATAAGAGC AACCAACAAT CATTACTGGT TTATTTTCCA GCGGATGAAT	4800
TTTGAATGAT AACCATTCTA ACACACTTTT CAAACCGGCA GGAATGGTAT GGTTATGTTT	4860
GGGCGTCGCA ATAATTACGC CGTCTGCTTG TAAATTTTG CGATTCATAT TCTGAATCAA	4920
GACACTGTTC GTTTGATCCT TACTTTGATT AAACATTGGA ATATCTTTAA TTTCTAAAT	4980
TTCCAAGTCG AACATTTTAT AAAACTCTTT GCCAATATAT TGCAGTAGTA ACCGATTATA	5040
AGAAGAATCA GCGTTTGATC CGACGATCCC AATTAATTTT ATTTTATTCA CTCCATTTCG	5100
TTAAATCGTT TCCCATGAGA AATTTTCTGC TTCTTTTTTA TTTTGCGCAT GCGCATGTTT	5160
CAGCTGTTCT GTAATTTCAA TAAATACACA AAAATCTGCA AACAGACCAG CTAGTTTATC	5220
TATTTGTTCT GAATCTTTTA ACGCATTGTT TTCATCAAAA GCTTGTAAGG AATGCGCTAA	5280
TAGAAATTCA GAGCTTGGCA TAATTCGTGC TTTCAATTCA GGAGCATCTA AAATTTGACG	5340
CAGATGCGCT TGTGCTCTAG AGGAACCTAA CGTGCCATAG GATGCACCAG TAATCATAAC	5400
TGGTTTATCA ACAAACGGAA AAATCCCGTA AGATAACCAA CTTAGGGCAT TCATCAAAGA	5460
AGCCGGCACC GCATGGTCAT ATTCAGGCGT ACTGATAATG ACGCCATCTG CTGCTTCAAT	5520
TTTTTCAGCA ACCGTTTTAA CTATTTCTGG CAAAACTTTG TCTTCTGGCT TGTAAACAT	5580
AGGAAAGTCA ACAATTTCAA CTAGTTCAAT ATCAGCCTCC TCGCAGAAAT AGTGTTGCAT	5640
AAATTGCAAA AGTTGACGAT TGGTTGATTG CTCTGAGTTC GTTCCTACTA AGCCGATAAT	5700
TTTTTTCATA ATGAGCTGTC CTTTCTTTTT CGTTTCTATG AATGGATAAT GTGAAGAGTA	5760
TCTGCTAACC CGTGCTATA TAAAATTGTC TGGTCTGTG TAATGATGAC TCCTTCAATC	5820
CCATCTAGTT GATTGAGGAT TTCAAGCGCT TCTGGAATTG GATAACCAA TAAACGCGTT	5880
GTCCAAATCT CACCATCCAC TGAAGCATCC GAGATAATTG TGAGGCTGGC CATTTCAATT	5940
TCCAATGGGT AACCGTTGT CGGATCCAAC AGATGGTGAT AGACTCGCCC AGCTTCGGTT	6000
AACTTCTTT CATAAATGCC CGACGTCACA ACCGATTGAT TAGCCACTTC GACAAGTAAG	6060
CTGATTGTTT CTCTGGATTC TTGCGGATTG CGAATACCAA TTCGCCATTT TTTGTTTTGA	6120
TGCGTAGAGG GcCCCAACGT CACAATATTG CCGCCCAGAT TAATCAAAGC AGACGTCACA	6180
TGGACTTCTT TCAAATAGGc AATTAGTAAA TCAGCGATAT AGCCTTTTGc TAAAGCCCCC	6240
AGATCTAATT TCATTCCtTC TTCTTCCAAA AAAACCGTTT GTTTTAAGGG ATTtAAGTGA	6300
ATCTTTTctG GATTAATTTT TTTCAAACAG GCTTTAATTT CCTCTTCTGA CGGGACACGG	6360
GCATCTTTAA AACCGATTCT CCATGTTTGA ACTAACGGGC CAATGGTAAC ATTCAAATGA	6420
CTAGCTGGAT CACAACTATG TTTTTCCCT AATGCAATTA ATTGATACAA CTCTGGATGA	6480
ACGCTGACTG GTTTTTGCCC AGCTTGTTGA TTGACAGCCA TCAGTTCCGA AGTTGAATCA	6540
TTTGCACTAA AGCGTTGTTC ATAAGTGATC AATCGTTGAT GCACTTCTTC TAGAATTTTT	6600

TCAGATTCTT CATGGTCTAC AAATACATCT ATTACTGTGC CCATCAATA AATAGTTTGA	6660
CTTTGTTGCA CCTCAAGACT TCCTCTCCAC GAAAGATTGT GTTTTATTTC ACATTTATTC	6720
ACAAATGAAT CACACGCCTT TATTTGATGC CAGTCAAGTC GTTTATATAT TTTTTATAT	6780
AAAATAACGA CCTTAAATTT TTCATTAAAG GTCGTAAGTT GTCTTGAAGT CTTTTAAAA	6840
ACGCTTTTTT TATTATATTT TGAAATAAA TATATAATTT TCTATATAAA CTAATTTCCC	6900
TTAGATAAAA TTGTTTCTTT CAATGTTTGA ATTGCTTCCA ACAAGATGCG ATTTTTATTG	6960
ATTTTTCAT ACGCGCGGGC ATCTTGTAGT AGTTCTAAAA TTTGCTCTTG CGGCTTTTC	7019

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32768 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

TATTTATTCT TTTAACGAAA CACACAATGA AGGTAAAGAG aCTCTtGGTG GTAAAGGGGC	60
AAACTTGGCA GAAATGACAC GCTTAGGTTT GCCGATTCCA CAAGGATTCA CGATTACAAC	120
CCGTTGCTGT ATGGACTATT TGGCAGATGC TACTTTCTTT GAAGAACACT TACAATCAGA	180
AATTTTGAAA GCGGTTAAAA ATCTAGAAAC TGAAACAGGC AAATCTTTTA CCGCGGACAA	240
TGAGATTCTC TTAGTTTCCG TTCGCAGCGG TGCTGccTtC TCGATGCCTG GTATGATGGA	300
TACTATTTTA AATTTAGGTT TAAATGATCA ACGTGTGAAA AAATTTGCGA CCTTAACCTC	360
TCCTGGTTTT GCTTTTGACT GTTATCGCCG ACTTATTCAA ATGTTTGGAG ACGTCGTCTA	420
TCACATTCCC AAAGAATTAT TTGACCAACA AAAAGAGCGC TTGGAACAAG AGCTCAACAA	480
AAAAATTACC GCTTTTCAAG AAGAAGATCA TTTTGCCTTG ATTGTCCGCT ACCAAGAAGT	540
TTTTGAACAA CATCAGGTGG TCTTCCCGCA AGACCCTGTC GCACAACTTT TTGAAGCAAT	600
CAAAGCTGTT TTTGATTCAT GGAATAATCA ACGTGCGGTG GTTTATCGAA ACTTACATCA	660
TATTGCACAT GATTTAGGTA CAGCTGTCAA CATTCAAGAA ATGGTTTTTCG GCAATCGTGG	720
ACTAGACAGT GGCAGTGGCG TTGTTTTTAC TAGAAATCCG GTCACAGGGG AAAATCAACT	780
GTTTGGAGAG TTTCTTTTGA ACGCGCAAGG AGAAGATGTA GTGGCTGGTA TTCGGACGCC	840
AGAACCCATT CGTCGTTTAC GTTTGACTAT GCCAAAAGTT TATCAAGATT TTCGCCATTA	900
CGCAGAACTT TTAGAATATC ATTATCGTGA CATGCAAGAT ATTGAATTTA CCATTGAAAA	960
TGAAAAATTA TATATTTTAC AAACAAGAAA CGGGAAACGA ACCGCCGCAG CTACGGTTAA	1020
AATTGCTTTA GACTTAGCCA AAGAAAATCG CATTACTAAA CAAGAAGCAC TGTTACGTGT	1080
GACACCTGAT ACGATTGATC AATTAATTCA TCCCGTGTTT GATCAAGAGA AACGACAACA	1140
CATGGAACGT TTAGCTATGG GACTTCCTGC TAGTCCTGGT GCAGCAAGTG GCCAAATTGT	1200

TTTTACAGCA	GAAAAAGCCA	AAGAACTGAC	CAATCTAGGA	AAGAAAGTCA	TTCTTGTTTCG	1260
TCAAGAAACT	TCACCAGAAG	ATATTGAAGG	CATGGTTGTA	AGCGAACAAAt	TGTGACCAGT	1320
CGTGGCGGAA	TGACTTCTCA	TGCCGCTGTG	GTCGCACGAG	GAATGGGAAC	TTGCTGTGTG	1380
ACTGGTTGTG	AAAGTTTAAC	CGTCAATGAA	GAAACCAAAC	AACTACACTG	CGGGCCACAG	1440
GTCATCTTAG	AAGGAACAAT	CATTTCCGTC	GATGGCTCAA	CGGGGGAAAT	TTATCTTGGC	1500
GAAATTCCTA	CAATTTCTGC	TGATAATAAC	GATGmCcTC	ArGAGCTCTT	GTCTTGGGCA	1560
GATGCGTaCG	CTGATtTaCT	GTTCGTGCCA	ATGCAGAAAC	TACaCAAGAT	TTAGAGACTG	1620
CTATCCGTTT	TGGCGCAGCG	GGGATTGGTT	TAGCGCGAAC	AGAGCATATG	TTTTTCGGTG	1680
AAGAGCGTGT	CTTGGAATG	CGTCGCCTGA	TTTTAGCAGA	GTCTGAAAAG	GAAGCAACCT	1740
ATGCGTTGGA	ACAACCTCTT	CATTTTCAAC	AGGAAGACTT	TTATCAAATG	CTGAAAGTAG	1800
TTCAAGACAA	GCCCATGGTG	GTCCGCTTGT	TAGATCCGCC	GATGCATGAG	TTTTTACCTC	1860
ATGAAAAGAA	TGATATTCAA	TTATTGGCGA	AgCAACTACA	ACgGTTCCCG	GTGACCATCG	1920
CCAAACAAAT	TGAACGGTTA	CAAGAAACCA	ATCCAATGTT	GGGTCATCGT	GGTTGTCGCT	1980
TAGGTGTGAC	CCAACCACAA	ATATACAAAA	TGCAAGTTAC	AGCCTTATTT	ACTAGTGCCA	2040
TTCGATTAGT	CAAAGAAGGA	ATTACTGTTT	ATCCAGAAGT	GATGATTCCT	CTAATCGCCG	2100
AAAAAGAAGA	ACTTCTCTAT	CTCAAACGCA	TTTTAAAAGA	AACGATTGAC	GGTCTCTTTG	2160
AAGAACATAA	AATGACTCCT	TTCCCTTACG	AAATTGGGAC	AATGATTGAA	CTACCTAGAG	2220
CTTGCTGTGAT	TGCGGATCAG	TTAGCAGAAG	AAGCAGACTT	TTTCAGCTTT	GGGACAAATG	2280
ATTTAACACA	AATGACCCTAT	GGTTTTTCTC	GTGATGATAT	CGGTAAATTC	ATTAATACGT	2340
ATCGGGAAAA	GAAATCATT	ACCCAAGATC	CGTTTCAATC	ATTAGATCAA	ACAGGTGTTG	2400
GTCAATTGAT	TCAATTAGCT	GTAGAGAAAG	CGCGTCGCAC	TAAACCAAAC	ATGAGTATTG	2460
GCGTATGTGG	AGAAGTCGGT	GGTGATCCGC	AATCCATCAC	ATTTTTCCAA	ACACTTGGCC	2520
TAAACTACAT	CTCTTGCTCA	CCTTATCGGG	TACCTATTGC	GAAACTCGCA	GCTGCCCAAG	2580
CCAAAATTTT	AACCGAACCG	GCTGTGACTG	AAGAACAAAT	GGTCCTTCTT	TAATTTAACG	2640
TGTGTAACAA	ATGATTCATC	CTTACAAGAT	AAAAAGTGCT	AGG <sub>r</sub> ACATCA	GTCGAAACGC	2700
TTGATGTTCC	TAGCaCTAAA	TTTTAAGATC	AATGATTGTA	CTCTGGTTAT	TTTCTTATAA	2760
ACGCGGCATC	TTTTGTCAAG	GGGTCTGTTA	CATCTGATAC	ATATCTTTCT	ACTTTCATAT	2820
GCAAATCATA	TTTTTCTCTA	AGTTCCGTAA	TACGGGTCAT	CATTGGAGAA	GCGTGGTGTT	2880
TGTCTAATGC	TTCCTGATCT	GTCCAACCTAT	CAATCAGCAG	AACTGTTTCT	TCATCGTCCA	2940
TTGGAAAAAA	GTATTCATAG	CGACTATTCC	CTTTTTCTGC	ACGAATCGCC	TCTACAATTC	3000
CGCTGGCAAT	CATTTCTTTG	GCAAATTCTC	TTGCACCTCC	ATTTGTCCCA	CTATAATAAA	3060
TATTTATAGT	AATCGCCATC	TGTCTTCCCC	CCTAAATTTA	CTTGCCTTTT	TACGTTCCCG	3120
ACTGGAATTG	AACCAGCGAC	CTACCGCTTA	GGAGGCGGtK	GCTCTATCCT	ACTGaGCTAC	3180



GAGACAATC	TGTTTTAATT	TTAGAAGTCT	TCACTACATC	TGTCAATAGA	CAAAAAGAAA	3240
GTATCCCGCA	TAAACAATTA	TGCGAGATAC	TTTCTTTTCAG	TGAaACACGT	CGTTCGTTTA	3300
TCTATTATTT	TTTTTGCCGC	CGATCAAGAA	TCATTTTTTC	TAAATAATAA	AGTCCTGTGC	3360
CAATAGCTAT	ATATTGCCCA	ATTTTATTAT	TGCGTGCAAC	TTTTTCTTGT	TTCGTTTTCG	3420
CCATTGTCCC	CATCACCTTT	CTAAAATAAT	TCTTTTACAA	TTTTTATTAT	AGTAGGTTCT	3480
GTCATTTTTT	GCGAATAAAC	ACACTTTTAA	AAATAATCTT	CTAAAATTGG	CATGCCTTTG	3540
ACTAACGTT	GGCctAAACG	TTGAGCTATC	GTTGCGTCTC	CTTGCAAGCG	TTCATAAAAA	3600
GATAACGTCT	CTGCTGAACG	ATAGCCCAAG	AATAGTTGCG	TCCACGTTTG	GATATCTGCT	3660
TTCAAAGCAG	CAGTCCCTTC	TTTTTCCGCA	GCACCTTTAG	TCACTGTTGC	TTTTCCTTGT	3720
TCGTCAATTG	TGATTGTCCA	AATGCCTTCA	TTCCATGGCC	CATAGGAATC	TTCAATCTCT	3780
AAAGAATACG	TTTCTTTTTT	TCCTGATTGA	AATGGATACT	TTTCCAGAAA	TGTTTGTAAC	3840
TCAACGATT	GTGCCATCAT	GTAAGGTAAA	ATTTTACAG	AAGCTGCTGG	CGTTGGCATT	3900
AAATCATTTA	AATCTTTTCC	AGCAAAACCA	TTAATCCAAT	GGAACGATTG	AACGGAACCA	3960
CTATGTGAGC	CAATAAACCC	AGCTAAAGCT	TTAAAGCTG	TATTAGTTAA	ATAGTTCCAT	4020
TCTACGATTT	CAAAGGTGCC	AGCAGCTATC	CGATAAATCA	CATAGCCTTC	TGCTTTCCCT	4080
TCTGACGAAT	AATAAATCGC	TTGATTATTT	GGTTTACTTG	CACGATTCAG	CGTATAATCC	4140
AGCCACCAAG	TTTCACGGAT	CACACCACCA	GAGTGCCTC	TTTGATTTTC	aAGGTATACG	4200
TCTTTAATGA	CCTCTTTACC	GTCAGCCCAA	GaAACTCgT	TAATGGTACC	TGGAACTCgT	4260
TTAAcACGTG	GCCaATCTTC	CgTTTTAAaT	GTATACTCTG	CTTGTTCaAA	TGTTTGTTCA	4320
TAGCCATATT	GACGATAAAA	TGGATATGAA	AACGGTGCTA	AGTAAGAAAG	GGCAACCTTT	4380
TGTTTCGCTA	AATCAGCCAA	CATTTCTTTC	ATAATAGCAG	AGATGCCGCC	TTCACCACGA	4440
TATTCTGGAT	AAGAAGCCAC	ATAGCCAATT	CCAGCCATGG	GATAGCGGAC	ACCATGGAAA	4500
TTCACTTGAA	AAGGTGTCGC	CATCACTTGA	CTTGTTAATT	GCTCATCAAT	TAGGAAGCCA	4560
TAGCTTTGCG	TATGTGACAA	CAATTTTTC	AAACGTTCTT	GcGTTCCGCC	GTGGGTTCCCT	4620
GATTAAAAGC	ATAGATAACA	AGATCAAACA	TCTCTTTCAT	TTCTTCTTTG	CCCATTTTTT	4680
TGACTCTTTT	AGTTGTCATT	TTCTTTCCTC	sTCGCACATT	GTATCAATTC	ATTTTATCAC	4740
AGAAACGATT	CTCTTCTAA	AAGCCGTTAT	TTGCTGTCAT	TTCTTTGAAC	CAACGACCGC	4800
TTTTCTTAAT	GGTCCGTTTC	GCATCATCAT	CTAAATCCAC	AGCGATGAAG	CCATAACGAT	4860
TTTTATAGGC	ATTGAGCCAA	GACCAGTTAT	CCATACACGT	CCACATATGA	TATCCTTGAA	4920
CATTACTACC	TTCTTGAATC	GCTTGGTGCA	CCCATTTC	GTGATCTTGC	ACGAATTCAA	4980
TCCGATAATC	ATCTTCAATG	ACTCCTTGTT	CATTGACAAA	GCGTTCTTCT	CctTCTACGC	5040
CCATGCCATT	TTCTGAAATG	TAGCAGCGAA	TGTTGCCATA	ATTTTCTTTT	AAATTCATTA	5100
ATGTATCATA	AATACCTTTT	TCGTAGATTT	CCCAGCCACG	GTAAGGATTC	ATTTTTTTGC	5160

CAGGCATATC GTAGACATCA TAAAAATCTT CTGGAAGTAA GCCTTCTGGT CGTGTTTCAA	5220
CAGGTGTTTC TTTTGCTTTG GCACGACGTG GTTGATAGTA GTTTACTCCT AATAAGTCCA	5280
CAGTATTTTC AGCAATTATT GCTAAATCTT CTGGCGTTGT TTCTGGCAAT AAGTCATTTG	5340
cTTTCGCCCCA AGCAACCATA TCTTCTGGAA AATGACCTTT AACAGCCGGG TCTAAAAACG	5400
AACGGTTAAA GAAACCATCG ACAAATTTAG CCGCATTTAC ATCGGCTTCG TTGTGCTCAT	5460
CACGAGGATA AGTCGGCGTT AAGTTTAAAA TAATGCCAAT TTCACCTGAT AAGTTCATAG	5520
AGTGATAGAC GGCAATCGCT TTGGCACTAG CTAAAGCTTC ATGATAGAGT ACTTGAATCC	5580
CTTCTTTTAG ATTCACTTTA TCAGGATAAT GCCAGCCATA GAGATAACCA CCTTCAACTG	5640
GTACAATTGG TTCATTGTGC GTAAACCATT TTTTCACGCG ATCGCCAAAC AATTCAAAAC	5700
AGGTTTTGGC AAAGTCAACA TACAAATCAA CTACTTCACG GCTTTCCCAA CCGCCTTTTT	5760
CTTGAAGTAC CATCGGCATA TCAAATGAT ACAAGTTCAT AAAGGGCTCA ATCCCGTGTT	5820
CTAATAAATC ATCAATGACT TGATTATAAA AATCAACTGC TGTTTGATTG ACTTCCCTG	5880
TTGTAGGATC TGGAAATAAA CGGCTCCATT GAATGGATGT TCGAAAAGAG TTATGTCCCG	5940
TTTCTTTTAT TAACTGAATA TCTTCTGGT ATTTTTTATA AAATTGTGAG GTCTTATCTG	6000
GACCGACTTG TTGAAAAAAT TTTTCTGGCG CTTCTTGATA CCAAAAATCC CAAATATTTT	6060
GACCTTTACC ATCACCTTCA AAAACGCCTT CTGTTTGAGG ACCGCTGGCT GCAGAACCCC	6120
ACCAAAAATT TTCTGGAAAT TGATATTCA TGTTATTGTC TCCTTTACGA TTTAATCATT	6180
AAAAAAGGA TTGAACCCTC CGTAAGAGTC CGGACGATTC AATCCTTCGA TTTTTTACTC	6240
GGCTGCTTCT GCGGCTGCTT CTGTGTTCAA GTATTGATCT GTTGCTACTT TTACGAAAGG	6300
CATGTAGATC AAAATAGAAA TCAATAAGTT AACAGCTGCT AAAACTCCGC CGGCATACT	6360
CTTCGTTGCT AAAAAATCCGC CAATAATTGG TGGTGTTACC CAAGGTGGCA TAAAGGTTGC	6420
AGCTGGTACT AAACCAGTTG CTGTCGCCAC ATACGCAACA GTTACAAGAA CCATTGGGAC	6480
TAAATGAAT GGGATAAACA TAATTGGGTT TAAAACAATT GGTAAACCGA ACATGACTGG	6540
TTCATTAATA TTAAATAGTC CTGGTGCTAA ACTTAAGTTT GTAATGACTT TGTATGGTTT	6600
GTTACGACGA CCTACGATAA AGATAGCAAT TAACAGACCT AATGTTGCCC CTGTTCCACC	6660
TAAGTTTACA AATGAATCAA AGAAAGGTTT GTTTACAATA TAAGGAATTG CTTTACCTGC	6720
TTCTAAAGCT TTAATGTTG CATCAATCGC TGGTACGTG ATTGTTTGCA TTAATGGGTC	6780
AACCATGTTG GCACCATGTA AACCAAAGAA CCATAAGAAT GGTGTAATGA AAGCTAACAA	6840
TAATGCTGAT GGGTAGCTGT TTGCAAGTCC CmTAAATGGT TGTTGAACTG CTTCATAGAA	6900
TGAAATAACG ATATTGTCAA TTCCTAAAGC AAATAGAATT GATGTAATTA AACTAAATAA	6960
ACTGATkGTA ATCATTGCTG GTAATAATGC CgCAAATGAT TTTGCTACTG CTGGTGGTAC	7020
ACCATCAGGC ATTTTGATAA CTAATTTAGG ATTGCCAGAT AAACGTTGGA ATAGTTCTCC	7080
TGAAATAAGT GCAATAATCA AAGCAATAAA TAATCCAGTT GCGCCCATGC CTTCTAAGCC	7140

GCCAACTGCG	AAGAATGAAG	CTACAGAAAC	TGTACCGGCT	GCAATGCCGT	CTTTTcCATA	7200
AGATTTAACT	AAATGTGTG	CTACTAAAAA	GGCTACTAAA	ACGGCAAAGA	TTGCGAATGT	7260
ACCGTTCCAA	ATGTTACCAC	CAAAGCTTTT	CCAAATTGGT	GCATCGTTAT	ACATCATTGG	7320
GAAAATTTTG	TTTAAAAAGT	CCTCATATCC	TGGAATTGGA	ATATTATTAA	TTAATACCGC	7380
CAAAGCCCCT	AAAATCATT	ATGGCATCGT	TACGACAAAC	GCGTCACGAA	TTGCTACCAA	7440
GTGCCGTTGC	GCTGCGATAA	CTGACGCTTT	CGGCATAAAG	TGTTTCTCCA	TCCATTGAAC	7500
AAATCCGTCC	ATTGTTATTC	CTCCTAATTA	ACGCTTTCTC	AAGCGCGTTT	TTCTTATAAG	7560
ACGCATTGAC	AACGAGATTA	CTCGTTATCA	ATACGGCGAT	AAAGATCGAT	AATTTCTTTT	7620
GCTAAATCTG	TAAATGCAAT	CGCAGTCATT	AAATGATCTT	GACTGTGCAC	TGTTAACAAT	7680
GTTACTTGCA	TGTTATTGCC	TTTTGCTTCT	TCTGTTAACA	TTCTGTTTG	CGAATGATGT	7740
GCATGTACAA	GTGATTCTTC	TGCATCTGCA	ATTTTTTTGT	CTGCTAATTC	AAAATCCCCC	7800
GATTTTGCTG	CGGCAATCGC	TTCCATTGCA	TCGCTTTTAG	CATTTCCGCC	ATACATGATT	7860
AAACCCATGA	TTGCTTCTAA	GTTCTGTTGT	CCGTCCATTA	CCTCTACCCC	ATTTTCTGCT	7920
CTAGTTTGAT	GATCCCATTA	ATGAAATTGC	TTGATCTAAA	ACTTTTTCGC	CATTCATCAT	7980
GCCATAGTCC	GCCATGTTAA	TTACATCTAA	AGGAATCCCT	TTGGTTGTA	ATTTTTGTTC	8040
AAATTGCCCT	TTCATGAAAC	GAAGTTGTGG	GCCTAATAAT	AAAACATTCA	CCTCTTTATT	8100
TTCCAAGTTT	GTATCTGCTT	CAGAAGCCGA	TACTGCAAAG	ATGTCTGCTT	CCATACCACG	8160
ATCTTCTGCT	GCTTTTTGCA	TTTTTGTTAC	TAATAAACTT	GTGCTCATTC	tGCGGAACAA	8220
ACTAACATAA	TTGTTTTTTT	TGCCATTTTC	CTTCATCTCC	TAAATTCTCT	TTGTTTTGCT	8280
TTACACCTTT	ATATAAGCAA	GAAGCGTGCC	AAGTTTAAAC	ACATTAAAAA	CCTTGATATG	8340
ACAACGTTTC	TTGATATAGA	AAGCGTGAC	ACACTAAAAA	TAAAACGATT	ACACACAGCT	8400
TTTTTGTTGT	CGTTTCCAAA	ATAATAGGGC	AACAAAAGAA	AGTGAGCGAT	AACCCGAAAC	8460
GACTCATCGC	TCACTTCTA	AAGCTGAAGA	GAATCAACTG	GACAAAAGCC	GCTTTTCCCC	8520
TTTAAGCAGT	TTCCATCATT	TTAAGCATTT	TATAAACTG	TTCTTTTTTA	AAATAACTTA	8580
AAAAGATACC	TAATACTGCA	TCATTCAGTT	TTTGTAACGC	CCCATAAGAA	ACCATTGTTT	8640
CTGTATAACG	CAGTTCACAG	CTTTTTTCAT	CAATTGGCTG	TAGCTCATAC	CGCACCGAGA	8700
AATCATTTTT	AGTTGTGGAT	GTTGAAATT	GATAGGTCTT	ATTTTCTACA	GCTTCTTCAA	8760
TAATAATTTT	CGCGCGACTA	TTTTTACTAT	ATCTTTTGAT	ATATCAAAC	TTACGTAAC	8820
GCTTTCTTGT	AAGAGTTTTG	CCTGTTTGCT	TACGGATATC	AAAAATAACA	GAATTCATTA	8880
ATTGCGTGTA	AAATTCATCT	GCTGAAATAG	TTAATGTTTT	CGTAATTTCC	ATTACCGTTT	8940
CACCTCATCC	TTTGTGCGCT	CTTCTTCAGC	AACTTTTTTT	CCTTTAAAAA	ACAGGTAACC	9000
ACCAAGTCCA	ATGAGAAAAA	TCCCCGTTGT	CAGACTTAAT	AAATCATAGT	TTAAGGAGCT	9060
GGGATTCAAC	GAATACAAAA	CAATCAAGAA	GCCAATACTA	CATAAAAAATA	ATCCATTCTT	9120

AATCATTTTC ATCCATCCTA TCTTTT TAGA CTACCGTTGC TTGGAAGTAT AAAACGCCAC	9180
CCCAaGGTTG TTCCTTAATC AGCTGtTTTG CAACTTGTGC AGCAGTTTCA GGGTCTCCTT	9240
CAACTTCAAA AAATAACTTA ATACCATTTT TTTCTTTAAA AGTAAATGTT TCATTTTCCC	9300
CTTGATATGC CTCTAATAAA GTAATAATTT CCTCTTGCTT CATTGCACTT GCAATCGAAA	9360
TCATGCCGCT CGCTCCTTTC TATACAAGTA ATTGCTGTTT TTATTGCTTA TTGCTCTCTT	9420
TCGCCTAATG CAGGATCCAT TTTGTTAGCA GCGATAACAA ACGGTGTCCA AATCAAGAAA	9480
GTGACTACCA TACATACCAA CGCTACGATT GGTGCGCGCC AGTCGGCACC TGTCGCTAAG	9540
AAAGATAATA GCAACGGCGG TGTCACCCAT GCGACTTGTT GAGAACTGG GCGGACTAGC	9600
CCAAGATACG TTGCAACCCA GCCAATTGTA GTAGCAACTA ATGGCGCAAT TACAAATGGA	9660
ATAAACATAA TTGGATtCAG AACAACTGGT AAACCAAACA TAATCGGTTC ATTGATATTA	9720
AAGATCCCAG GCCCTAATGA TAATTTACCA ACTGTTAAT AATCTGCTCG TTTGGAGAAT	9780
AACAAGATTG CAATAATTAG GACGATTGTA CCGCCTGAAC CACCAAACCA AGCAAAAGCA	9840
TCAAACGAAC CACGGACCCA CATGTAGGCT TTACCGTCGT CAATAGCGGC TAAAACCGCA	9900
GCTGTTCTCG TTTTCTCTC ATAACCTTTT TGGAAAATAT CAATATTGAT TAATTGTGCT	9960
TGTCCCCAAA TTCCTTCAAG AACTGGTGCC AAAACGTTTG GCCCATGAAT ACCAAAGAAC	10020
CAGAAAATTT GAACAAAAAC CGTTACGATT AACACAGCGC CTAGTCCTTG TGATAACCCA	10080
AGGAACGGTT CCGCGATATA TTTTGAATT AAGTCAATTA ATAATTGACC GCCTGAAAGT	10140
TTTGACACCG TAAAGTTAAT AATTGCTACG ACATACAATG CGATTGTTGC TGGTAAAATT	10200
GCCGCAAACG CTTTTGAAAC TGCTGGTGGT ACAGAATCAG GCATTTTAT TGTAAATCC	10260
GCTTGCAATTA ATTTACAAAA AATAATGGTT GATAAAGCGC CCATAATCAT CACCGTAAAG	10320
AAAGCGTTAC CATTTAGATG GTTGAGATTT AACCAACCCC ATGGATTGAC AGTCAATTGA	10380
TTATTAGCAA TTGACGCACC AATTTCTGCC CCATTAATTG CGTTGGCAAT CTTTTCTGGC	10440
ACTTTTACAG CTAATTCAGC TGTGTTATTT GGTGCAAAAG CAACACCTGA AATTAAGGTA	10500
GCTAAACTGA CAATCCCACC AGCTAGCTCA TTCACATTGT ATGCTTTCGC TAAGTTATAG	10560
CCCCAAGAAA AAGCGAAAT CAAACCAGCG ATAsCcAATG TACCATTCCA AACAAAACCA	10620
TTAATCCCAA TGATGACATT CAATACTGCA AAAATACCTT TGTCTGCTTC CAAACTTGCC	10680
GCATAAGTTG AGACAAACTC TTGGGAAGA TCTCGAATAA TTGCGTTAAT CATTACGGCG	10740
ACTGATCCAG CCATTGTCGC AGGTAATGTA CCAATAAAAG CATCCCTTAG TGCCACCAGA	10800
TGTTTCTGTG cACCAATTTT TGCGGCAGTT GGTAAGATAT ATTTTCCAT CCATGCCATT	10860
AATCCGTTCA TTTTATTCC TCCTGTATCC ATTTATATTT ATTAAAAAGG CAGTAGCCTC	10920
ATTAAACGAA TGAATCGAGC GTACAAAACA AAACCTTTGG TAACATGCAC TATCCTTATT	10980
TGCTATTAC TTTTCCTAAA CGTTTTAATT CCTTATCCCA TTAATGAGAT TGCTTGATCT	11040
AAACTTTTT CGCCATTCAT CATGCCATAA TCTGCCATGT TAATTACATC TAAAGGAATC	11100

CCTTTTGGTT GTAATTTTGT TTCAAATTGC CCTTTCATGA AAQGAACCTG TGGACCTAAA	11160
AGTAAAACAT TCACCTCTTT ATTTTCCAAG TTTGTATCTG CTTCAGAAGC CGATACTGCA	11220
AAGATGTCTG CTTCCATGCC ACGATCTTCT GCTGCTTTTT GCATTTTTGT TACTAATAAA	11280
CTCGTGCTCA TTCCTGCGGA ACAAACTAAC ATAATTGTTT TTTTGCCAT TTTTCTTCA	11340
TCCCCTTAAA AGTTATGACG TTATACACTT TATTTATAGC AATAATTGTG CCAAGTTTTT	11400
TCCTCACTAA AATACTGATA TAACAGCTTT TGTCATCGAA AAATGCGTGC ACACACTAAA	11460
AGAAAACGCT TACACACATG AAATTGTGTG TAACTTACAA AAAGAAAACA TCAAGAAAAA	11520
AGAGAGCGTT TTCTCATTTT TTGATGTTTT TTTTATTTAT TATTGTTTAC TTTTAATTTT	11580
ATTTTCTAAA AACATTTGGG TAACATGCGC AACTTCCGCT TCTGGAATCG TCACACGATA	11640
CGCCCGTTCA ATTAACATAA AATTCGTGCG TACGACATCG ATTTCTAAAC GATGTGTTTT	11700
AACATATTCC GCTAAGTTTG GAAAATTTCT TGTTTCTTCT CCTTTAATTA AAGCATCTAC	11760
CAGAAAAGCC AAATGAATCA TGATTCCGGC ATCCACCCCT GGTCAACAA TAATATGCAT	11820
GTCTGTTTGT ATTTGATGAA CTGTTTTTTG TAACATGAGA ATTAATTTTT GTAAGGAGTC	11880
CACAGAAGTA ATGACGCCAC TTAGTGAATG AACCAATTGTA TCAATCGCCA CTTCATCACT	11940
AGCAATCCGT TTCAGTACAT TTAATTTTTT ATCATCAAAA ATATCATAGA CGGAAAAGAA	12000
GGGAATATTT TGATATTCAA CATCCaCAGT TCCCGCAATG GCCTTAATTT CATATTCTTC	12060
CATCAAGTTG TCGATATGTT TTTTAAATGT TTCTCGTTCA ATAAATTGCA TTTGAATCAG	12120
TTCCACTTTT GTTTCATCAA TAACTGGTAA GATTCGTTGG TATAGTTTTG CGGCAACCCC	12180
TTCTCCAGTA AAGCACGTCA CGATTACTGC TTTTTTTACG TTTTGACGTT TTTGCAGACT	12240
AGAACGAAAT TGTTCTCTCA CAACGCTTTC AAAAGAAAGT TGAATATTTT GGTAAATGTC	12300
CtCTAAACTt CGTCCCACGC TTGCCATTCT AATAGCTTCT AAGACAATCA TTGTGCTCGT	12360
CATTGTAATC GCCTTTGTTT GGATACCTGT TTCTTCAAAC AACATATTTT CAAACGAATT	12420
TAATGAACCC ATATCCGTCA GTAATaAAT CCCATTATTT AACGATTCTT TTTGCGTAAT	12480
TACTTGGTTT CGTAATTGTT cATACATTGT TTGAACTTCC ATCGTTAACG GCATATTCAT	12540
CGCTGTTCCG ATTGACGTTT CAAAAGTTC TTGGGCGGTT TTCAACATAC TTGAGGCAGT	12600
AGAATCCCCa TGCATTAAGA CGAACACACC AACATTATTT TCTGGAGTTC TTTCTTCTTT	12660
ACCGACATTA ATAGATAAAA ACATACTAAT AAAGCCAATT TCATCGAAAG GAATCTCTAT	12720
GTTGTTCTCT TCTTCAATAA TAGAAGATAA ATCTAAGGCT ACTTGAAATT CTTTTTTCAT	12780
ATTTTTACGG ACATTATTTA AATCAGGATG GACAATCACG TGACCTTCTT TGACACGTTT	12840
AATCGTACTT TGTAAATGAA GCGCAAAGGC aAAGCGCgCa GTTTCGTTGT ACGTGTGTTT	12900
CaAACGCTCT TCTGCGaTAT CATaAAGACG ATTCGTTAAG TCCCAAATAG CGTCTGGTAA	12960
TAACTCCTTA TGTAAAGAAT GTTGGGTAA TTCTnTCGACA TAATTTTTTAA AATACGCATT	13020
CATATCTTTG GAAATCAATT CTTCTAAATC AACATTTTCT AAGTCAGCTT CGGACAAGTC	13080

AGCGACTTTT TCTyCGATTT CaTTGTAAAC TTGCATGTCT CTAGCTGGAT CTTGTGACCA	13140
AACAATTTTG TTGGTGCCTG GCTGGAAAGA TAAATAATTA CTTTTACTAT CGACAAATTG	13200
ATCCAAGCGC TCAGGTACTT CTTTAATTTT TAATAATCCT TTTTGGACAG GTAACGACAA	13260
ATCCCGTTTA CGGATAACCA AATGTTCTAA ATTGTGCGTC CGATAATAAA GGAATGACTT	13320
AGCACAACT AATTTTAAGT CCCGCTTCAC CTGACCAATA TTTCTTCTG CATCGTATAA	13380
CATAAAGCC AAAACTGCTT CGCGTTCGAT ATCAATTTGT TGGCCCAAAC GTGTTGCCTC	13440
TTGCTGAATA AATAACGAAA TGATGGCATA ACGCTCATCT AGCGTCCGTG CATCTAAAGA	13500
TGGTAATGTA ATCGACATCG GAATCCGCCG ATTAAGAGTG GTTAAAAAGT TTTCTGAGGA	13560
TTCCGTTGTC GCACCAATAA TTTGAACAGA GGCTTCATAG GTTTGACCAC TTTCTCCTAA	13620
TGGACGGTAC ACGCCTTTAT CAATAAACGT AAAGAGCATT TCTTGACCTT CTGGCGGTAA	13680
GCGATGAATT TCATCTAAAA AGAGAATCCC GCCATCTGCT TTTGCCATTA AGCCTGGACT	13740
ATCTTCATTC GCACCAGTAT AGGCACCTTT TTTTACGCCA AAGATATGAC CAAATAATAA	13800
TTGTGGGTTT TGTGCATAAT CAGCACAATT AAAAGAGACA AAAGGGGCAT CTTCATCTAA	13860
GGTTTTTGAA TCAATGGCAA AACGGTACAT ACATTGAGCA AAAAGAGATT TCCCTGTTCC	13920
TGTTTTCCCA AAAATAATCG TATGCAAACC TCTTGGTGGA TATAAAATAG CCGCCTTAGC	13980
CTGCTGAATA CTCACTTTTA AAGAATCATT GGCACCTACT AATGTGTCAA AACTAACAGA	14040
GGTGGACTCA TCTACATGGA CTTGTTCTCC TTCATTGATT GTTGGCACTG AATAGATAAC	14100
AGGACGTCCA GTACTTCTAG TAATTTTCTT TTCTTTGAAT AATTCATTTA AATAACGGCT	14160
GGCATTACTG CGATCAATCA TTAAGTGTTC TGCTACCGCT CCTGCCGTTA AGCCAGTTTC	14220
ATTTTTTTGA AGTAACGCTA AGATTTCTTC CTTCCGAGAT TTCATGGAAT CCTCCTCGAT	14280
ACAACGTTCT GCTTGArGCT CGTAAAGCGT TATTTCTTCG CTTTACGAGC yCCTTTGTTT	14340
TTCGTGTTTT TCTTTTTGTT TTTCTTCTTT yTGC GTGTTT CTTGTGGTTC TTTGACAACA	14400
GAATTTACTG TTTCTGGTTT TCTTTCTTTC GTACTCTTTT TTTCTGAATA CGTATCTTTG	14460
ACTACGTTCT TTTTTTCTAG TTCATCTTTG GCTTTTGGCT GTTCATGAAG TGCACCATCA	14520
TATAAAAAGA TTTCTGATA AGGAATGGCC AATTGTTTCA TCAAACGTTG ATAATCCCTT	14580
TTGGTGCCAT CGTTGATAAA CGTAATCACT GCCCCTTGTG CACCCATCCG GCCAACGCGA	14640
CCAGAACGAT GAATATAGCT CTCTTCGGTT AAAGGAACAT CTACGTTTAC GACGTAAGGG	14700
ACCCAGTGA AATCTAAGCC CCGTGCCGCA ATATCCGTGG TTAATAACAA CTTCACTCGT	14760
TCTGCTTTAA ATTGATCGAT CGCCAATTTA CGCAACTGTT TACTTTGATC TGAAGCCAGA	14820
CCAATAGCCG GAACATTCTC ATAACTAAT TTTTCTTCTG CTGCACCTAA ATCAGCAACT	14880
TGATTGAAAA AGACCATGCT TCTGAATGCT TCTGTATGGG CCAAACGACG TAAGTAATCA	14940
GCTTTTTTTC TTGGTGCTAA CCGAATAAAG TAGTGGGCGA CTTGGCCTGC TGAAGTATCT	15000
TCTTCCGTCA CATCAATCAC CTGTAATTTT TGGCTTAAAG ATTGCGCTTG ATTCACGACA	15060

CGATCAGCAG TCGCTGAATA AAAACAAGT TGGTATTCGG TTGGTGTGTG TGTCAAAATT 15120  
TGCTTGGTTA AACTTAGCTC TTCCTCTTGG AAAAGTTGAT CAACTTCATC CATTACGATT 15180  
GTTTTTAACA ACTGAGCTTT GACTTTTTTA TTCTTCATTA ACTCTAAAAT TCGACCTGGT 15240  
GTTCCGATTA ACACTTCGGG ACGTTTTTTT AACTTGTCAA TCTGGCGTGA AACATTTGCC 15300  
CCGCCGATTA ACGTTTGAAC CTTTAATTGA AGTGGCTTAG CCCAAGTCCG AGCAACTTCA 15360  
GCAATTTGCA TAGCTAATTC TTGAGACGGA GCAATAATCA ACAATTGATT TCCCTGTCCT 15420  
TTTTCAACAG TTAAGAGTAA AGGCmGmTG TAAGCAAGTG TTTTTCmGT TCcTGTCGGC 15480  
GAAATTCCTA AAACATTTTC TCCTtCGCGA AGGGGCGTAA AACTTTTTTC TTGAATTGTT 15540  
GATGGTTCGT TAAAGCCAGA AGCTTGCCAC TGTTCTTGCC AAGCCTCTGG AAGTTGTTCT 15600  
AAAAATCTCA TATTGTTCTT TTCCTATCGC TTAATTCTTA TTTATCAGCA TCAAAAATAA 15660  
TGCCGGCACT ATGACGTAAA TCAGTACTA CTTTATTGAC ATTTGAGCA AGCTCAACCC 15720  
ATTCTTCATA CTTGGTGCCC CAATACTTAT CTGTTGGGT ATTTAAACT GCTGCAAAAT 15780  
CACGGGCTTC TTCAATCATT GGATTATCGG CTAATTGAAT ATCTAAGACA TCACGTTCTT 15840  
CATGTGCGCG ATCATGGTAC TCAGCCGAAG AAATAGCATT GACACTATCT AAAATCAAGG 15900  
TCCCGTTTTC AAAGTAAAT TCGGATGGCG CAAAAGAATC ACTAATTTTC CCAGTGTTTA 15960  
ACGTACGTC AACTGATCA TATCGTAAAA TAATTGTTCC TAAGCCATCG ACACCTGTCG 16020  
CAATTTTACT AGCAAAATAA TGGCTTTCGT TTGGTACGCC AAACCAGCCT AAGGCTGCAT 16080  
AAACTAAATA aACGCCTAAA TCCGCTAAGG CTCCCCAGA AAAATGAGgA rAAAAtATTT 16140  
GGCTCTTTC CTTCTAATAC TTGATCATAG CGAGAAGAAT ATTCATATA CGTAAAATTC 16200  
GCACCTAAAA TTTGGTTTTT TAATGGCAAT AATTCAGCAA TTTTTGGAA ACTTTGCTCA 16260  
TGAATATTAC GGGCCGCTTC AAAGAAATAC ACACGATTTT TATTCGCTAA TTCAATAATT 16320  
TCTGCCATCT CATCAGGCGT TGAAAAAGCT GGTTTTTCAA CAATTACATT TTTTTCGCT 16380  
AAAATTGCTT GTTTCGCTTG TTCAAATGT AATGAATTTG GTGAGGCAAT ATAAACCGTA 16440  
TCCATATGTG CAATACCAA AAACGTCTCT AAATCAATTG CATATCCAC ATCCCCATAG 16500  
CGAGAACCAA ATTCTTGTGC CGTTGCTAGT TTCCGAGAAT AAACAGCCGT TAAATCATAC 16560  
GCACCCGTTT CTAAAGCTGC TGCAACAAAT TGATGACTAA TCCAATTGGT ACCAATTACG 16620  
CCTAAATGTA TCATAATCAA AAACCTCCAT TTTTGTTTA TTTACTCTCT TTATAATAGT 16680  
AATCTCTCGT ACAAAGTCA CGTAATTACC CTGCCTTAAA GCGCTAGGAG TTCTTCAAGT 16740  
TATTTATTAG GATCATCAAA TGCCTGTAAT AAATCAAAGG TTACCTGTGA CTTAAGTTCA 16800  
ATCGTTGCTG GTGTTACGTG GCAATTGTAA GCAAGGACTT GCACACCTTG ATCAATTGCA 16860  
GCACCCACAT TTTCAGCAA GGCGGGTTC ATCTCTGGT GAAGCGTTC TTGCTTAATT 16920  
TCTTCAAATT GGACCACAAA CAAATGTAA CAGCGGTAAC CCGCTTTAGT TGCAGCCATT 16980  
AGTTCAGTGA CATGCTTCAA GCCACGAAGT GTTGGCGCGT CAGGAAAAGC ACCGATGCCT 17040

TTATTTTCTA AAGTCATGCC TTTCACCTCG AAAAAAGCTT GTTCGTCCGC ATCCGTTTCC	17100
ACTAGAAAAT CAAATTTTGA ATGTGCAAAT CGTTGCTCAC GTTTAACAGT TTGGATCGTA	17160
CCAACATAAC CAGGAAGCAC AATTTGACCA TTTTTTAACG CTTCAATTGAC TAATGTATTCT	17220
GGTACTTGGC TATCAATATT AATCCAAAAA GAACCTTTTT TTAAGTCAAT CAAATCGTAA	17280
TCTGTTTTTC TTTTAGGCCA TGGTTGATAA CTAAGCGCCA CAACAGCACC TGGTAAAAAA	17340
ACTTCTTTCC CTCTGCCAGT ATTTTGTACA TGTGTGATTA CTTCTTCATT CGTTTCCATT	17400
AAACGACAGT GTGCAATAAA CCGATTGGGT CGTTCGATAA AATAAGCAAG TTGGACATTC	17460
GGGTACTTCA TACACAGACT TCCTTTCTTA CACACTATAC CATAAAAAAA CAAGGAAAAG	17520
TATAGACCAA AACCAACAAT TTTACTATCT TGTGTCAATC ATTAAATATT TGTCATTTCT	17580
CAGCGATTGA GTCGAATATT TTAGACAAGT TCTATATAAT AAAGGTAACG AAGACGAATA	17640
AATGAAAATG AAGGTTAGGA GGAATTTCTT TGAAAAACA TTGGACTTTC TTGGTTGGTG	17700
CATCTTTACT TACTGGATTA GCAGGCTCTA TCCTCTGGAA AAAAAATAAA CAGCTCCCTT	17760
CCCCACAGGA GCTTCAACAC AACTATACTG GGCATTGGTG GTTGTGGAC GTGAAGAAAG	17820
CCACGCAACA CACCTTAGAA ATCACCTCAT CTTTCAATGT TTTCATTGAT GGCAACAAA	17880
CGCATGCCTT GCTTGTGAA TTAACCCCTA AACGCTTAGT TATCTTAGAC AAGTTAGGT	17940
ACCATTTAAT CATTCAATGT GAAAATGGTG TCCCTGCGAC GCTTTACGAT GAAGCCGACG	18000
ACGAAACCTA CACACTAGAA CGGGCCATTT ACAATTAATA AATCACCAC GACACATAAA	18060
ATGAAGTATG TGCGGTGGGT GATTTATTTT TTTACATCCC TTGCAAAAGC ATTCAGAAAG	18120
TTTTCTTACA ACTTGAATAT TATTAAATAG TCAATGCGC CTTTTTCTGA TAACTTAGG	18180
ATGATGTTTT TTTGAAAGGA TGAGCCCCTA TGACACCAA TCGCGAAGAC TATTTAAAA	18240
TAATTTTTGA ATTAGGTGGC GACGAAGTTA AAGTGAATAA TAAACAAATT GTTTCTGGAC	18300
TCGATGTTTC GGCAGCTTCG GTTAGTGAGA TGATTCAAA GTTAGTAAAA GAAGATTGG	18360
TTGAGCATTC TCCTTATCaA GGGGTACAAT TAACTGAAAA AGGCTTAAAA AAAGCGAGTA	18420
CGTTAATTCG CAAACACCGA ATCTGGGAAG TCTTTTGTAGT AGAGCACTTA AATTACACTT	18480
GGAATGATGT GCACGAAGAG GCAGAAGTTT TAGAACATGT TACTTCACAG ACGCTTGTGA	18540
ACCGTTTAGC GGATTATTTA AATCATCCAG AATTTGTGCC ACACGGTGGT GTTATTTCCG	18600
AAGATAATCA ACCCATTCAT GAGGAGAAAC GCCAAACGTT AACAGACTAC CCTGTTGGCA	18660
CAAAAATTCG GATTGCACGT GTCTTAGACG AAAAGAATT ACTGGATTAT TTAGTTTCCA	18720
TTGATTTAAA TATTCAAGAA GAATATACGA TTAAAGAAAT TGCTGCATAT GAAGGACCGA	18780
TCACCATTTA TAATGAAAAC AAAGAATTAT CCGTCAGCTT TAAAGCAGCA AACACAATTT	18840
TTGTTGAGCC GTTGATTAGA GAAAGTGAGG AAAACTAATG AATGAAAAAG TAGCATTATT	18900
TACTATTTTC GGGGGGACAG GTGATCTTGC CAAACGTAAA TTGTATCCTT CCCTGTTCCG	18960
CTTATATAAA AAAGGATTAT TAGCTGAGCG TTTTGCGGTG ATTGGTACCG CACGTCGGGA	19020



ATGGACAGAC GAATACTATC GTGAAATCGT CCGTGAAACC ATCAAGACT TAAATCCCAC	19080
GGCCGAAGAG GCAACTGAAT TTTCTAGTCA TTTCTATTAC CAATCACATA ATGTGAATGA	19140
TACGGAACAT TATAATACCT TAAAAGAATT ATCTGATCGC CTAAATGAAC AATACCACTT	19200
AGAAGGCAAT CATGTTTACT ATTTAGCTAT GGCCCCCTCAA TTCTTTGGAA CCATTGTTAA	19260
TCACTTGAAA TCACAACATA TTATTTTCGA AGAAGGCTTT GATCGTCTAA TTATCGAAAA	19320
GCCATTTGGT TCTGATTATG AATCTGCCTA CGAATTAAAT GAAGAAATTC GTGCTGCATT	19380
TCCTGAACAA GATATTTTCC GAATTGACCA TTAAGTAAAGC AAAGAAATGA TCCAAAATAT	19440
TTCAGCCATT CGTTTTGCCA ATAATATTTT TGAATCACAA TGGAACAATC GTTACATTGA	19500
TAATGTCCAA ATCaCCTTCG CTGAAAGCTT AGGTGTCGAG GATCGTGGCG GCTATTATGA	19560
CCACAGTGGC GCACTAAAAG ATATGGTTCA AAATCATATT TTACAAGTCG TTGCATTGCT	19620
TTCTATGGAG CCTCCTGTTG CCTTTTCTGA AAAAGAAATT CGGACTGAAA AAGTAAAAGC	19680
TTTACGTGCC ATCCGCCTTT ATTCTGAAGA AGAAGCCCTT CAAAATTTCTG TGCGTGGACA	19740
ATACGGCGAA GGTGAGCTAG CTGATCAAAC ATTTGAGGCC TATCGTGATG AACCAAATGT	19800
GGCAGAAACT TCATCCACTG AAACCTTTGT TGCTGGTAAA TTTTAAATTG ATAATTCCG	19860
TTGGTCTGGC GTCCCTTTCT ATGTTTCGAC AGGTAAACGA TTAAGTGAAG AAGGCACACG	19920
CATTAACATC GTCTTTAAAC AAGTACCTAT TAATGTCTTC AAAGATGATA CGTGTGAAGA	19980
ATGCGATAAA ACCGATTTAC CACCGAATGT TTTAACTATC TATATCCAAC CAACTGAAGG	20040
ATTTTCATTA ACCTTAAATG GAAAAGAAAT CGGTCAAGGG TTCAATACAA CTCCAGTGAA	20100
ACTTGATTTC CGTCAAAGTG CTGAAATGAC TGAAAATAGT CCAGAAGCAT ACGAAAACT	20160
TCTATTAGAC TGCCTAAATG GCGATAGTAC AAACCTCTCT CATTTGGGACG AAGTTGCCCCA	20220
ATCTTGCGCG ATTGTTGACA TTATTCGACA TGCTTGGGAT AAAACAGACG TTTCTTTCCC	20280
TAATTATGCC GCTGGCACAA TGGGCCCCGA AGCTGCCTTC GACTTATTAG AAAAAGATGG	20340
CTTTACTTGG GAATGGCAAC CAGATAACTG GTACCGTGAC CGCGGTCAAC TTGATCAATA	20400
AATAAGCCAA AAAAGATCCG CTAGTTAACG ACTAGCGGAT CTTTTTCTT GTTTATTCTGA	20460
ATCCTTCAGC AACCGCTTCT GGATAATTTT CTTCAACAAT GTGTGCACAA CGGCCaCAA	20520
GTGTTGGTAA TTTTTCATCT GAGCCGACAT CTTGACGAAC GGAACGACAA CGGTCACACG	20580
TTTCGCCTTC TGCTTTTTCA ACTAAAATAG CCATGTCTTC AAATTGAACA GCTTCACTAG	20640
GTGCTACTTC TTTTGATACT TCAAAGTCGG AAACAATTAG TAATTGAGCA ATATCTGCAT	20700
CAACAGCTGT CATTAAATTGA CGAATTTGTT CATTTGGATA AACGGTCACT TTGGCTTCTA	20760
ACGATTTACC GATTAATTTA GAATGACGAG CTTCTTCCAA CGCTTTTAAG ACATTGTCAC	20820
GGAAATCCAT AAAGGCTGCC CATGTATCCA TCAATCTTTC TTCATTAGTA AACGTTTCGT	20880
AACCTGGGAA TTCAGCTAAT TGCACATACT CTTCTTCTTC TTGTAAGAAA CTCCAAATTT	20940
CTTCCGCTGT ATGTGGAATA ATTGGTGTCA ATAATTTTGT TAACGAAACC AATGTTTGAT	21000

AGAAAACAGT CTGCATACAA CGACGTTGAT AATCGTTTTC AGCTTCAATA TAAACGACAT	21060
CTTTCGCAAA GTCTAAATAG AAAGAAGATA AATCTACAGT TAAGAAGTTC ATAAGTGTCC	21120
GATAAATATG CATGAAATTA TATTTTTCAT AACCATTTTC ACGGATTCT TGAATGACTT	21180
GATTTAAACG AACCGTCATG TATTTATCTA CAGAACGTAA ATCGGCATAA GCTACAGTAT	21240
GCTCAGCTGG ATTAAAATCA CTCGTATTCTG CTAACAAGAA ACGCATAGTA TTACGGATT	21300
TCCGGTACAC TTCAGATACT TGGTTTAAGA TATCCATTGA GACACGCACA TCGGCTTCCG	21360
CGTCA\CACT ACTTACCCAC AAGCGTAAGA TATCTGCCCC CATTTGATTG ATAACTTTTT	21420
CAGGTAAGAT AGTATTGCCT AACGATTAC TCATCTTACG GCCTTCGCCA TCGAGAACCA	21480
TCCCTTGAGA AATAATTGAT TTATAAGGCG CTAATCCATT GATAGCGACA CTTGTTGTAA	21540
TACTTGAGTT AAACCAACCA CGATATTGGT CAGAACCTTC TAAATACATA TCTGCTGGGA	21600
AAGTTAACTC TTCTCTTCT CGTAAACTC CTTTATGTGA AGAACCTGAG TCAAACCAAA	21660
CGTCCATGAT ATCTGTTTCT TTGGTAAACT CGCCATTGGA TGAACCTGGA TGTGTAAAGC	21720
CTGCTGGTAG TAACTCTTTT GcTTCACGCA TAAACCAGAT ATTTGATCCA TGTCTGCAA	21780
ATAAATTCGC TACATGTTCA ATTGTTTCTG GCGTAATGAT GGCTTCCCCA TTTTCAGCAT	21840
AGAAGATCGG TAATGGTACA CCCCACGCTC TTTGGCGAGA GATTACCCAA TCACCACGGT	21900
CACGAATCAT ATTATATAAA CGTGTTTTTT CCAAGGAAT TAGCCAGTCA ACTTTCTCAA	21960
CTTCATCTAA AATATCTTGA CGGAATTTAG AGATTGAAGC AAACCATTGT GCGTTCAC	22020
GGTAGATAAC TGGTTTTTTG GTACGCCAAT CATGTGGATA ACTATGCGTA AAGAAATCTA	22080
ATTTTAATAA CGCGCCTTTT TCTTCTAATA ACTCAGTAAT CATTGGATTG GCTTTGTCAT	22140
AGAAAATTCC TTCAAAACCA GGTGCTTCAT CTGTAAAGAC CCCACGGCTG TCAACAGGAG	22200
AAATAACAGG TAAATCATAT TTACGGCTAA CGATATAGTC ATCTTCACCA TGTCTGGTG	22260
CGGTATGAAC TAAGCCAGTC CCGGCATCTA GCGTCACGTG GTCGCCTAgC ATAACAAGTG	22320
aTGTTCGATC ATAGAATGGA TGTGAGCTG TCATACGATC TAATTTTCA CCAGAAAATT	22380
CTCGCAAGAC TTCTACTTCT TsCCAACCAA TCGCTTCTTT AACTGTTGTT AATAAATCTT	22440
TAGCAATTAC AAATTaCGAC CGTCTGCCTT CACTTCAACA TACGTAAAAT CAGGATTAAC	22500
TGAAATACCT AAGTTGGCTG GTAATGTCCA AGGGGTTGTC GTCCAGATGA CGAAGGCTGT	22560
CTCGTTATCT AAAAGACCTT TACCATCGGC CACATTAAAA GCTACGTAAA TAGAAGGTGA	22620
TTTTACATCT TTGTATTCAA TTTCTGCTTC TGCTAAAGAA GATTCACTTG AAGGAGACCA	22680
ATAAATTGGT TTAAACCTT TATAAATATA GCCTTTTTCT GCCATCTTAC CAAAAACACG	22740
AATTTCTGCT GCTTCATACT CTGGATCTAA CGTAATGTAT GGATGTTCCC AATCGCCTGA	22800
CACACCTAAA CGTTTAAAT CGTTACGTTG TTTATCGACT TGTGATAAGG CATACTCTTT	22860
ACATTTTTCG CGATACTCAG CGACAGTCAT TTCTTTACGT TTAACCCCTT TATTGGtTAA	22920
CACTTGCTCa ATTGGTAAAC CATGAGtATC CCAACCAGGc ACATAAGGAG AACGGAAACC	22980

AGACATTGAT	TTTGaACGaA	TAATAATATC	TTTACTGATT	TTATTCAAAG	AATGTCCTAA	23040
ATGAATATTT	CCGTTTGCAT	ACGGAGGGCC	ATCATGTAAA	ACAAAGGTG	GTTTTCCCTC	23100
GTTTAATTTT	TGACGTTGTT	CATATAAGCC	TTTTTCTTCC	CAGTCTTTTT	GCCaTTCTGC	23160
TTCACGGTTT	GGCAAGTTCC	CACGCATTGG	AAAAGCTGTT	TTTCCTAATT	GCAATGTTTC	23220
TTTCATTTTC	ATATGCATTT	GCCTCGCTTT	CATTTTATCG	ATTTAATTGT	GTTAGTTCAT	23280
ACTAAAGATG	AACGATAGCT	GATAAAAGTA	GATAAAAAAA	GAGTACTCCa	TCCCaAGGGA	23340
CGAAGTAcTC	GTGTTACCAC	CCAAATTCCa	AAGTAGAGTT	GCCCCTACTT	TCTCTGAAAT	23400
TGATAACGGT	AATTATCCGG	CAGTTCCTAC	TCAAGTGTTT	AGAACTGCAT	CTAGGATAGA	23460
GGATCTGTTT	AATGCCAGTG	TCTGTCAAAC	TTTCACTATC	ATTGACTCGC	TGCTAAGAAT	23520
AAACATTAAA	CGTTCTGTTC	TATCTATTTT	GATTCTTCTT	CAATTGTTTC	TTCAGTGACT	23580
TCAATTACAG	GCTTTTTTTC	TACTACAGCA	TCTGTCTTCT	CTTCAGAGTT	TACAACAGTT	23640
TCATTTTCGT	TGTCAAGGTC	TTGTTTCATCA	AGAATTTCCCT	TAAGTCTGCT	ATGTTTGTCT	23700
CCGACATAAC	TAGAAAAAGG	CTTTAAAATC	TCTTCCCACT	CTTCACTCTT	TACTTGcTCT	23760
AATTGAGTTT	CAAGCATTAA	GCTTAAGCGT	TGaTGGAAGA	CACGTGTTTT	CTTCTTCAAG	23820
TCTTCTGTTT	CCCCAGCTAA	TTGACGGGCA	CGTTCAATTG	CTTCTGCTAA	GATTTGTGTT	23880
GATTTTCGCTT	CAGCATCTGC	AATCATTGCA	TTTGATTTAC	GTTCACTTC	TACCAAGGTT	23940
TCTTTGGCTT	GGTTATCTGC	CGAAGTGATA	ATCATTCTCTG	ATTCTTTATT	GGCACTTGAT	24000
TTCACTTTAT	CAGCGGTATC	TtGTGCAACA	ATAATTGATT	GATTCAATGC	ATCTTTCAAT	24060
TCATtGAAGT	ATtGTAATTT	TTCTtCTGCG	TGTTTTAAtG	AwTTCTCAAG	TTCACGATTT	24120
TTTTGTAAATG	CAtCTTCATA	ATCACGTGTG	ACTTGATCTA	AAAAGTCATC	TACATCGTCT	24180
TGGTTATAGC	CTCTCATTTT	TGTTGAGAAG	tCTTTGTTTT	GAATAwCTAA	tGGAGTTAAT	24240
GCCATATTCC	TTCACCTCTT	TGGATTATTT	TTGTAAAACC	CCTTATGAAC	ACATTCTCAT	24300
ACTAATCCTG	AAATAGCAAA	TATCTTACTT	TTTTTATTAT	ACACAACCTC	GTCTGAATGA	24360
ACAGCCAAAA	TGGTTTTCTT	AAACAAAAGT	TCGAGATTAT	TTTCTCAATA	CCCCTAAGAG	24420
GACACGAAAT	TTTTCTTTCT	TGGTTTTTCC	TTCTAATTCT	TGAATTTGAA	GGCGCCCAAA	24480
TCCCCGTATC	GAAACAATAT	CTAATAAGTC	TAAAGCAAAA	TCTGGTCGTG	TCGTTTCTGT	24540
CCAATTGACT	TTTACTTTAC	CAGATTTCGAT	CAATTGTTTT	GATCGCTGCC	TTGAAATATT	24600
AAATACGGCG	GAAATCACAC	TATCCAGCCG	CAAAGAGCTC	ATAGTCGTTT	GCTCTTCCTG	24660
CCAACCATCC	TTCGGTATCA	AAATTTGTGT	ATAATCCCGT	TCTTCTAAGC	GAACAGTAAT	24720
TTTACCAATT	TTGTCGACTT	GGTTAACAAC	GAAGCCAGCT	ACTTCTTGTT	CTATAAAAAAC	24780
CTGCCACTTA	TCACCATCGG	AAATAATATC	TCCAAAGGCT	TCCcTACGAA	CCCCAACATT	24840
CATCAaGGTG	CCTAAAATTT	TCCCaTGaGA	TAaTGTCGCA	AaTTTAGATG	GATAAATaAT	24900
kTCAAATAAT	ACAACACCAA	AATCTGTCTC	TTCTGGTATG	TAATACGCTG	GATAAATCAA	24960

GCAGCGTTTT	CGTTCTGCTT	GCTCATAACC	ACCATAAAAA	GTAAACGATA	AATCACTGTT	25020
TTGTCGAATC	AACGTTTCCA	AAATATACGC	CTGTCTAGGA	TCAAGAAAGT	CTGTCAAATA	25080
AGGCGCGTAT	TGCATCTCCA	CTTGCTCTAG	CCAATCCCCG	ACAGTGTCAA	TAAAAGGATG	25140
TTCATCTTTG	CGGAAATGTT	GATACACATT	TGCGTTCATA	GATAAATCAT	CCTTTCTAAA	25200
AGTTATTAAT	AAACAATAGC	TCGTAAAATG	ACTGCTAAGC	CACCCGCTGC	TAGGTTAAGG	25260
ACAATAATCC	CTACCATTAC	GTTAAATCCA	ACCATTCCAA	TCCTTAAATT	CAAACGGTCA	25320
AACAACTTA	AGTATGGTTC	ACAAATTCTA	GAGAGCAAAC	GCCCAGAAAGC	AGAATCATAT	25380
GCTCCTGGAA	ACCATGAAAG	TAGCGCGTAA	ATCACTAGTA	ATCCTGAATA	GATTTGAACG	25440
CCTTTGTATA	ATAGAGTAAT	CAGTGTGCG	ATAATAACAC	CTCCCTAAGG	CTTAGAAATC	25500
AAAGAATTGT	TTGTTGGCTA	AAGATTGTGC	TGTCGCACTA	TCAATTTCCA	TATTCGCTGG	25560
TGTGCATAAG	AAGATTTTCT	TGCCAACGCG	CTGAATATCT	CCGTCTAAAG	CATACACAGT	25620
ACCAGTTAAG	AAATCAACGA	TGCGTCGTGC	TTGGTCTTCT	TCTACTAATG	TAAAATTAAC	25680
TAGCACAGCT	TCTTCAGCAA	ATAAATGTTT	CGCAATGCTC	ATTGCTTCTG	AATACACACG	25740
TGGTTCAATA	ATAGTAATTT	TATGTGTTTT	TGCATTGGCG	GTGCTTCTTT	GTGCTCGTCT	25800
TGTATTCGTA	GTCGCTGATT	GATTAGAACT	ACGCATAGAA	ACCACTTTTT	TCTCTGTGTA	25860
AGCAGCTTGT	TGTGTGTCTC	TTTCTTGATG	CGCTTCAGCA	GAACGATAGC	GTGCAGAAGG	25920
TTTTTCAACT	ACAGGCTCAC	TTTGTGTGTG	ATAAGCTGGC	TGACTGTAAC	TTTCCTGTTG	25980
TTGAACAGGG	CGTTGTGGTT	TAGGCCGAGC	TGCACGACGC	GGTGGTTCAT	TAAGTCTTTT	26040
GCGCTCTTCA	TATTCTTCGT	AGTTATCATA	ATAATCATCT	TCTTCGCCAG	ATAAGCCAAA	26100
AAAGCTCGAT	AACGCATCCT	TGTTAAAAAT	TGACATAAAT	TCCTCTCCTT	TACGCATCTC	26160
TGAATAACGC	TGTTCCGATT	CGTATAAAAG	TTGCACCTTC	TTCAACTGCG	ATTGGGAAAT	26220
CGTtACTCAT	TCCCATGCTA	AGTTCTGTAC	AAGGAGCAAA	AGATAAATGT	TTCTCATTTA	26280
CGGTCATTTG	CAGCTCTTTT	AATtCTCTAA	ATGTTGCATG	TAATACTTCT	TCACTTGCCC	26340
CAAAAGGTGC	CATAGTCATT	AGACCGACGA	TTTGAATTTT	GTCCAAATCA	GCAAGTTCTT	26400
GAATGAAAGG	TAAGACGTCT	TCTTTACGAA	AGCCGTGCTT	GCTTTCTTCC	CCAGTTATAT	26460
TCACTTCAAC	AAAACAACGA	ATAACTTTCG	TTGCTCTTTT	TTGAATTTCT	TTGGCTAGAC	26520
TCAGACTATC	TAATGCATGA	AAATAATCTA	TCTCATTAAT	AATTAATTTT	ACCTTCCTAC	26580
GTTGTAAATT	ACCAATTAAA	TGCCACTGAA	TATCATTAAT	TTTTTCCAAT	CCAGCTTTCT	26640
TTTCTAGTAA	CTTATCTACT	CGGTTTTTCAG	CTAAATTCTG	GACACCAAGC	TGAACAAGCT	26700
CTGCTGCTTC	CTCAATTCCT	ACTGATTTAG	TGACGCCAAT	CATTGTCACA	TCAGCAACAG	26760
AGCGCTGTGA	GCTTTTGCAT	GCTAGCTGAA	GTTGATTATT	AATATTTGTT	AAGTTTTTCT	26820
CTATCATTTT	TAATCGATGC	TCCTTATCGT	TTTCTGCGGA	AAAATGGTGG	CGTACTTAAT	26880
TCGTATCAT	TGTGATTTGG	TTCTTCACGG	TGGAATGTGT	CAAATTTCTT	TTTGTCTACT	26940

TGTTCTAATG AAGATTCATC AACTTTTGGG CGTGTATTTT GTTCTCTGCG AATATCCCAA	27000
TCGCCAAAAG CACTTGCTTC TTCTTGATGAT TTTGGTTGAT CCATTTCCAC AGATTGAGTT	27060
GTTTGTGCA TTGGTTGAAC CGCTTGCTCTT GTTTGACGGT GCGGTTTTCT ATCCTTTTTT	27120
GATTCATCAA TACCTGTAGC GATTACTGTT ACACGAATTT CGTCGCCTAA ATCTTCGTTA	27180
ATTGATGTAC CTAGGATGAT GTTCACATCA CCTGAAGCAG CATTGTGAAC GATATCTGAA	27240
GCATCTTGCG CTTCAAACAA GGTCATATCT AAGCCACCAG TAATGTTTAA TAAGACTTGT	27300
TCCGCACCAT CGATTGATGT TTCTAACAAT GGTGAAGAAA TAGCTTTTTT CGTTGCTTCA	27360
ATTACACGTT CTTGCGCGCT CGCTACACCA ATTCCCATTAA AAGCTGTTCC TTGGTTTTCC	27420
ATCACTGTTT TCACGTCAGC AAAGTCCAAG TTTACGTAAC CTGGTGCAGT GaTTAAATCT	27480
GAAATCCCTT GAACACCTTG aCGTAATACA TTATCAGCTT CTCTAAATGC TTCAAGCATT	27540
GGCGTTTTCT TGTCAACGAC TTCTAATAAG CGGTTATTTG AGATAATTAA TAGTGTATCA	27600
ACGTTTTCTT TTAATAAGGC AATTCCTTCA GCanGCAAAc GACCACGTTT TGGACCTTCA	27660
AAACTAAATG GACGAGTTAC TACACCAACT GTTAAAGCGC CTAATCTTT AGCGATTTTT	27720
GCAACTACTG GCGCAGCACC TGTACCAGTT CCGCCACCCA TACCAGCAGT AATGAAAATC	27780
ATATCCGCGC CTTGTAATGA TTCTGAAATC ACTTGTTTAC TTTCTTCTGc AGCTTTTTTG	27840
CCAACTTCAG GTTGTGAACC GGCACCTAAA CCACGAGTGT ATTTAGGGCC TAATTGAATC	27900
ACTGTTTCTG CTTTTGAATG TTTTAATGCT TGAACGCTG TATTGGCTGT GATAAATTC	27960
ACGCCTTTAA CGTTTTCTTC GATCATACGG TTTACAGCGT TTCCGCCGCC GCCACCTACA	28020
CCGATAACTT TAATGACTGC ACCGTTGTTA ATGTTATTGT CTAATGAAAA TTCCATGATA	28080
TCGTATCCTC CTGTTTGTGT GATTAGTCGA AAATGTTTGA GAAGAAGTCT TTGATTTTGC	28140
CTGTCACTTT TTCACCAGAC TCACGTTCTG TGAATCTTTC GTATTCTTCT TGAGGCTGTT	28200
CAGCATATGT GTCATAGCGA ACTTCTTGTT GGACAGCGAC TGATTGTGCT GGTTCGATT	28260
TTTACCAGG GATTGCATAT TTGGCAATGT GATAAATATC GTTTAGTTGC GCTGAATATT	28320
CCACAATGCT AATGACATTA GCAAAGACTG GGTTACGTAA GCCCATGTGA TTTGGTACGT	28380
ATAGTTTTAC GTTGGCTTCA AAAATCTCTT GCGCCAAATC GACAATACCT GGCATACTAG	28440
CGGCGCCACC AGtTAACACA ACGCCTCCAG GTAATTCmAA CGCATCAATT TCATCCAGAA	28500
CTTCTTTTGA TTTTCTCAAA ATTTGTTCCA CACGCGCTTC AATAATTTCT GAAAGGTAAC	28560
GTTTCGTCAC TCTTACAGGC TCTGACTTAC CGATAACATC AACAGGAAAC TCTTCGTTTG	28620
CTGAAGTTCT TTCTGGATAA GCATCCCCAT AATTAATTTT TAAGGCTTCG GCGTTGTAA	28680
AAGAAGTATT CAATACGATT GAGATATCTT TCGTGATGAA CTCGCCGCCT TCTTGGTTAA	28740
CATGAGTAAA TTTCAATTGT TTGTCGTGGg ATAACAGATG TCGTTGtTTG aCGGCCACCC	28800
aTATCaATGa CAATkGtTCC aAAGCTTTTT TCGcCATCTG tTaAAATTGT TTCCGTTAAA	28860
GCAAGCGGTG tAATCACTAA TTCGtTAATT CCTAAgCCAG CTTTTTCAAC ACATTTACGG	28920

ATGTTATGGA TGaTTGTTTT TGGtCCAGTA TAGACAACGC CAAACATCTC CATGCGAACG	28980
CCAAGCATT CACGTGGGTC TTTGATGCCT TCAAAGCCGT CGACGGTAAA GTCTTGCGGT	29040
aAGATTGCAA CAATTTGACG TTCAGGAGGA GTTGAACGAA CCAAAGCTGC AGATGCGACG	29100
TTGCGCACAT CCTCATCTGT AATTTCTTTT GATTCACTGC TAACAGCTAT CATTCTTGA	29160
CAGCTTTCTA CTTCTAATAA ATTCGCTGGT AATCCAACAT TTACGCTTTT AATTTGAATG	29220
CCGGCTTTTT CTTCGGCTTG TCGAACGGCC CGTTGGATTG CTTGAACTGT TTGATCAATA	29280
TCGACGACAA TCCCTCGATT CAAACCGTCT GATTTTGCAT TTCCAACGCC GATGATGTTT	29340
ATTTGACCTT CAATATATT CAGCGACAACT ACTTTGACAG AAGTTGTACC AATATCGAGG	29400
CCTACATACA TTCCTGTTTT TGCCATGAAC GGAATTCCTC CTCCTATTAA TTCCAATATC	29460
TAAGCGGACA AAGGCGTCCG TTAACTATC TCTTTTTAG TATTATTCAC TTCTTCCAAT	29520
TTTACCATAA AAATGGGGGT TTTAGAAAGC AGATTATACG AATTTGTTTCG TTAATTTTG	29580
AAACATTTTT GTAAAAAAT TATTCTGCTT GTTGCACTTC TTGATTTCT CCATTTGTCG	29640
GCGTACCGTT ATTCGCTTCA TCCGTTGCTA CGCCTGCATT TGGGTCCACA ACTTCTTGAT	29700
TTTCAACCGC TGAATCTTCG GAAACTTCAC TACCCGTTT TCACTTTCT TTTTCATACG	29760
GATAAGAAAA AATGCCGACT TCCATATCAA CGATTCCAGG TTTGTCCATT TGCTCGGCTA	29820
CTTGTGAATA ATAAGCCATT TTTTCACTTA AATCAGAAAT ATTAACAATC ACTCGATTT	29880
CATCATTCAT ATAGACATTA ATCAAGCTT TATTGGTTTT GGATGGTGCA TATTTAATTT	29940
CTGAAATTCC TTGTTTAATT TCTTGAGGTA ATTTGTTGTA AGAGGCCATT AATTCAGGGA	30000
TTAATTTATC CTCTTTGAAA TTCTCAAAAA TTGGTTTGCC ACTTTCAGCA GCTTTGGTGG	30060
TTTCCGCCAA AGTTTACCA TTCTCTAAAA TCGGATGATA GCCGCCTTTA GTCGAGCAA	30120
GAGCGACAAT TTGATACTCT TGAATATCAA TTTTAAATGA ATTAATACCA TTAACTTAA	30180
TGTTGCGCTT TTTTATAATT GGAACTTTT TCTGAATATT TGCAGAAAA TAATTTCTAT	30240
TTGAATACTG CTCCATAAG CCACTGCCCG TTTCAAGTTT TGACTGCTGA ATAATCGCTT	30300
GGCTCTCCAC ACTTTTGTTA CCTGAACTG TTACTTCACT CAATCGACTA AGAGGCGAAA	30360
CAAAATACAA TGCCACTAAA ATGGCTGTCC CTAAACAAGT TAAAATAAAG GCTAAACGCC	30420
TATATAAGCG CTGTTCCTGT GTTTTCTTAA CGTTCGGTAA ACGATCTGCA AAATGCTCCT	30480
CTTCTTCCAC TGTTTCTTGT GGAAGATGTT GTTGAGGCTC TTCTTGAATT GTTCTTGTA	30540
AAGTGATGGT GACTTCTGCT TGTTCTTTTT CACCTTTTGA AGCAGCTTCT TCCGCTTGCT	30600
TCTTCAATA TTCTTGTTT TGTTTTTGCC AAGGTGTCAG TTCAGGTTGA GACTCTTCCG	30660
ATTTTTTTGG AGGTAATGGT TTTCTTTTTT TGAATATATC GTTCTCGTTA CTAATCTTCC	30720
ACACCTCCTA TCTCAGACTA ACGTTTTTAC AACTTGATAC AAACGATCAC TTGCATCTGG	30780
GATGCCTTCC CCTTTGGACG CCGTTGCCAT CTGTTGTCTT TTTTCATTGT TTAATAAAAT	30840
ATCATCTATT GCCGCAACTA AACGAGCGCC GGTCATTCT GCATCTGGAA TCATTTCAAC	30900

AGCACCAACT TTTACTAAGC TTTGCGCGTT TTTCGTTTGA TGATCATTCG TGACGTAAGG	30960
ACTCGGAATT AAAATTGCTG GTAATCCTAA TGCGGTGAAC TCCGCAATTG AGGTTGCTCC	31020
TGCACGTCCC ACCATTAAAT CGGTATTAGC CATCACTTCA ACCATTTTAT CAATATACGG	31080
TTGCACACTT ATGTTTGTCA ATTTCTTTTC AGAAAACTTT AGACTTTCTT GAAGTTCCTG	31140
ATAATATCTT TCTCCTGAAG CGTATAAAAC TTGATACTCT CTCTCTTCAA ACAAAGGAAA	31200
GGCCTGTTC AAGCCTTGAT TGATTTTCAA GGCGCCACGA CTTCCGCCAA ATAAAACCAC	31260
CGTTTTCTTT GCTGGATCCA AGCCAAACTC TGATAAAATA GCTGACTTTT CCACCGTTAC	31320
GACTTCTTGT CCCCGCGGGT TCCCTGTAA AATTGTTTTT TCTTCGGGA AAAAAGTAGC	31380
GACATCCGGA AAACAAATCG CAATTTTATC CACATAACGA CTTAAAAATT TATTTGTCAT	31440
TCCGGGGATA CTATTTTGTT CATGAATAAT TGTCGGTATT TTTAATTGGT GGGCCGCATA	31500
AACCACTGCA CCAGAAACAT AACCGCTGT ACCAATGACC ACATCTGGTT GAAATTCACG	31560
AATAATTTTT TTGCTTTAT TAATGCTTGT TAAAAATAAA TAAATTGTTT TAAAATTTTG	31620
AGGACTTAGC GAACGCTTGA ATCCTTGAAT TTTAATCGTT TTAAATGGAA TTTTCGCCTT	31680
TGGCACAATT TGACTTTCTA GACCATTTCTC TGTTCCACA TACAAGAATT CTGTTGCAGG	31740
TGCTTCTTTT TTGACGTGTT CGACGAAGGA TAACGCAGGA TAGATATGCC CACCTGTGCC	31800
ACCACCAGTT ACTAGTATCT TCATTTTCTC ACATTTCACT TTCTTTTAAT TGTTGAACAG	31860
CTTGCAATAA GGCTTCCCCG CGTACTTCAA AATTCGGGTA TTGGTCCCAA CTTGCGCAAG	31920
CAGGTGATAG TAAAATAGTA TCATCTTTTT CCGAATAATC AAAGGCAATG GTAACCGCCG	31980
TTTGAACATT TTCAGCAAAT AAAATTGTTT CAATGTTTCGC TTTTITAGCA GCTTCCGCCA	32040
ATTTTTCTTT GGTTCCTCCA AACAAAACAA TTGCTTTGAG TCCCAACAGG GCAGGAACCA	32100
ATTCATCAAA TGAGTTACCG CGATCCAAGC CACCCGCAAG TAAAAGTAGC TTTTGGTTGT	32160
CAAACCCACT TAACGCCATC TCTGTAGCTA AAATATTGGT TGCTTTTGAA TCGTTATAAA	32220
AACGTCTTTG CTGAAGTTCT CCAACAAATT GCGTTCGATG AGGAACGCCT GAAAAGTTTG	32280
TCAAAGTTTG TCTAATCTGA GCATTCGATA CATTTTTTAA TTTAGCTACA CAAATCGCTG	32340
CGAGTGCATT TTCAATATTG TGAAGTCCAG GAATCCCTAG CTCATCGGCG GGCATAATAT	32400
ATTCTTCATT GAAATATAAT TTCCCATCTA AAAGATAAGC CCCTTCTACT GCTTCTTTTCG	32460
TTGAAAAAGG CAATACGTTG GCAGCTGTGG TTTTGGCTAA CGTTTGAAGC TCTACTTGAT	32520
TCCAATTTAA AATCAAGGTG TCCTCTGCGG TCATGTTTTT TTGAATGGCC CATTTTGCAG	32580
CAACATATTC TTTCCGCGAA CCATGATAAT CCAAGTGTGC CTCAAAAATA TTTGTAATTA	32640
CTGCAATTTG TGGGTGAAAC GTCTCAATTC CCATTAACTG AAAAGTAGAA AGTTCCATCA	32700
CAAGATTATC CTTGGCCGTT GCTTCTTGAG CCACCGTACT CGCTGGAAAA CCAATATTTT	32760
CCGCCAAA	32768

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6259 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

GGCTGGATTT TCCTAGTAAG AGCTTGTTGG AmCAGCTTAA TAAAGACACA AGCCTTTTTT	60
GACGTCGTCG CTCAGGATTT TGCTAAAATG GCCCAGTTTA CGTATCATTT ACAAGAAGCA	120
AAAAAAGTCA CAGAGTATAC AAATTCTTTA AGATTAAAGG ATTTACAAGG AAAATAAGCT	180
GAAACAACA TAAAAATAGT GTATAATTGA AAAAATTAAT CTGCGAGAGA AAGGGAACCT	240
TTTTTGATTG TAGATCACCG CAGTTGTTTG TTGCCTAGTC TCTAAGAAAT AACCTAAAA	300
TTATTGATAT AAAGGATGAA CAAATGAAAA AAGAAGAAAT GCAAATGCGT AATACACGTC	360
GTCAAAAATC AGGAAAAAAT AATAAAAAGA AAGTAATTAT TACTTCTTTG GTTGGACTAG	420
CTCTGGTTGC TGGGGGCAGT TATGTTTATT TTCAAAGTCA CTTTTmCCA ACCACAAAAG	480
TAAATGGAGT TTCTGTAGGC TGGTTAAATG TAAATGCTGC AGAAGAAAAA TTAGCGCAAG	540
TTAATCAAAC CGAAGAAGTT GTGGTTCAAA CGGGGACAAA AGAAGAAAAA ATTCAACTTC	600
CTAAAAATA CCAATTGGAT CAAAAATTTT TAAAAGACCA TTTACACAGT AGCAAGGTGA	660
AGCTACCGTT AAACGAGGCA TTCAAAAAAG AACTAGAAGC CAAATTAGCA ACTTTGAGTT	720
TTCCAGAGGG GAAACCAAGC AAAAATGCGA GTATCCGTCG AGGCAATGGC ACTTTTGAAA	780
TTGTTCCCGA AGAACAAGGC ACAGTAGTGG ACACACAGCG CTTAAACCAG CAGATTATTG	840
CGGATGTTGA AGCGGGAAAA GGCAACTATC AATATAATGC CAAAGATTTT TATAAGCCC	900
CTGAAATTAC AAAAGAGGAT CAAACGTAA AGGCAACATT GACAACGCTC AATAACAAGT	960
TAAATAAAC AATTACAGTT GATATTAATG GTGAAAAAGT AGCCTTTGAT AAAACACAAA	1020
TTCAAAACGT GCTGAATGAT GATGGCACAA TCAACAAAGA AAAACTAACT ACTTGGGTGA	1080
CACAATTAGA AACAACATAT GGTTCGCTA ATCAACCAGT TTTATTTACA GATGTTACG	1140
GCACGACACG TCGTTTTAAA AACAACGGAA GTTATGGCTG GTCGATTGAT GGGGCCAAAA	1200
CGCAAGAACT ACTAGTAAAC GCGCTGAATA GCCAAGAACA AACGAATGCA ATCACTGCTC	1260
CGTTGGTTGG TGATACCAA GAAAATAGTA AAATTGCCAA TAATTACATT GAAATTGATT	1320
TAAAAGATCA AAAAATGTAT TGTTCATTG ATGGCAAAAA AATAGTCACC ACAGATGTCA	1380
TTACTGGCAG ATATAACAAA GGAACCGCAA CAGTACCAGG ATTCCATACA ATTTTATATC	1440
GGACAACCGA TGTGAATTTA GAAGTCAAA TGCTTGATGG TTCTCGATAC AGTGTGCCAG	1500
TAAAATATTG GATGCCGTTA TTAAGTCAAG GGGGCGTTGT CACACAAATC GGGATTCATG	1560
ACTCCGACCA TAAATTGGAT AAGTATGGCG ATAAAGAAGC CTTTAAACC GATGCTGGTA	1620
GTAATGGCTG TATCAATACG CCAGGAACAG AAGTTTCAAA AATCTTTGAT GTATCCTATG	1680
ACGGAATGCC GGTAATTATT TATGGACATA TCTATGATGA TGCACCAGGT GAATTTGATA	1740



AACCTGTAGA TTACGGCGGA GAGGTATATA AAGAGATATT AGGCAAGACC GTACTTTGGT	1800
CCAGCTGTGT ATCTGAAPAG TAAATATTGG AGACGTCAGC GTAACGACTG ACGTCCCCAA	1860
TTTTTTTATTG ACAAGCGTCA TACCCCGCTC TATGATACAA ACAGAACAAG CAATTTAAAT	1920
TGGGGGAACG AAGATGCCAT TTGTACATGT AGAATTAATT GAAGGCCGCA CAGAAGAACA	1980
GTTAACTAAT ATGGTCAPAG ATATTACAGA AGCTGTGTCA AAAACGCTG GTGCACCAAA	2040
AGAGAATATC CaTGTGATTG TCAATGaaTT GAAAAaACC GCTATGCACA AGGCGGCGAA	2100
TGGAAAAAAT AAGTTTTGAG AGAATTTCTG TTGAATCACT TGCTTTTTCT GAAAATAAGT	2160
GGTTAAATAA GTGCATAAT GATTGATGAG AAAGACAAC TGAATTTGCCC GCGATCTATT	2220
AGAGAGGAAA ATCTTGGCTG AAPATTTTCC AGATTGGCAC AAGGGATGAC CACTTTGAA	2280
ACCTTTGTTG AAACAAAGGC GGTGCGACCG TTAACACGCT AGAGGAACAC AGATTATTTG	2340
TGTTCAAAAT TAGGTGGAAC CACGATATTT TCGTCTAGT ATTCGTTAGA ATACTAGGAC	2400
TTTTTGTTTT TTAACAAGGG GAACAGGTGG CTGTTTCGTT TCATTGTGAA AAACAGGAAC	2460
ACAATGAAAC GCTTCAAATC AAAAAAATTA AAAGGAGAGA ATTTTATTAT GTCTATTCAC	2520
ATTACTTTTC CAGATGGCCG TGTAAACCG TTTGATTCTG GAATTACAAC ATTTGATGTT	2580
GCTAAAAGTA TTAGCAACAG TTTAGCCAA AAAGCTTTAG CTGGTAAATT CAATGGTGTT	2640
TTAATCGATT TAGATCGTCC TATCGTAGAA GATGGTTGCG TTGAAATCGT GACACCTGAT	2700
CATGAAGATG CTTTAGGAAT TTTACGTCAT TCATCAGCTC ATTTAATGGC TAATGCCTTA	2760
CGCCGTCTTT TCCCTAACAT TAAATTTGGC GTAGGTCTG CGATTGATTC TGGGTCTAT	2820
TATGATACAG ATAATGGAGA ATCCCTGTG ACAGCGGAAG ATTTACCTGC AATTGAAGCC	2880
GAAATGATGA AAATTGTGA GGAAATAAC CCAATCGTTC GTAAAGAAAT CTCACGTGCA	2940
GAAGCGTTAG AATTATTTGC TGATGATCCT TACAAAGTTG AATTAATTAC AGATTTGCCA	3000
GAAGATGAAA TCATCACTGT CTATGATCAA GGCGATTTG TTGATTTATG TCGTGGTGTT	3060
CACGTCCCTT CAACAGGACG GATTCAAGTC TTTAAATTAC TTTCACTAGC TGGTGCTTAT	3120
TGGCGCGGAA ACTCTGACAA TCATATGATG CAACGGATTT ATGGCACTGC CTTTTTGAT	3180
AAAAAAGATT TAAAGAGTT TATCAAAATG CGCGAGGAAG CCAAAGAACG TGACCACCGT	3240
AAATTAGGaa AgAATTAGAT TTATTTATGG TTTACAAGA AGTTGGTTCA GGGTTACCTT	3300
TCTGGTTACC AAAAGGCGCA ACCATTGCTC GTACAATTGA ACGTTATATT GTGGACAAAG	3360
AAATTAGCTT AGGTTACCAA CATGTGTATA CACCAATTAT GGCAGATGTG GAATTATACA	3420
AAACATCTGG TCACTGGGAT CATTACCATG AAGATATGTT CCCACCAATG GATATGGGTG	3480
ATGGCGAAAT GCTGGTATTA CGTCCAATGA ACTGTCCACA CCATATGATG GTTTATAAAA	3540
ATGACATTCA TAGTTACCGC GAATTGCCAA TTCGAATCGC TGAATTAGGG ATGATGCACC	3600
GCTATGAPAA ATCTGGCGCA TTATCAGGGT TACAACGTGT TCGTGAAATG ACTTTAAACG	3660
ATGGCCATAC TTTGTTCTCT CCTGACCAAA TTAAAGACGA ATTTAAACGT ACTTTGGAGT	3720

TAATGGTGGC AGTCTATGCT GACTTTAACA TTACGGATTA TCGTTTCCGC TTAAGCTATC	3780
GTGATCCAAA TAATACAGAC AAATATTTTG ATGATGATGC GATGTGGGAA AAAGCGCAAG	3840
CGATGTTAAA AGCTGCCATG GATGAATTAG AATTAGATTA CTTTGAAGCA GAAGGCGAAG	3900
CTGCCTTTTA CGGTCCGAAG TTAGATGTTT AAGTAAAAAC AGCTTTAGGA ATGGAAGAAA	3960
CATTATCAAC CATCCAATTA GACTTCTTAT TACCAGAACG TTTTGAAGTA ACTTATGTTG	4020
GCGAAGATGG TGAAAATACA CATCGCCCAG TTGTTATCCA CCGTGGTATT GTCTCAACAA	4080
TGGAACGATT TGTGGCTTAC TTAACAGAAG TTTACAAAGG CGCTTTCCTT ACTTGTTAG	4140
CACCAATTCA AGCAACTATT ATCCCAGTTT CTGTAGAAGC GCATTCTGAG TATGCTTATG	4200
AAATCAAAGA ACGTTTACAA GCACAAGGCT TACGTGTTGA AGTCGATGAT CGTAACGAAA	4260
AAATGGGCTA CAAAATTCGG GCCTCTCAA CACAAAAAGT ACCTTATCAA TTAGTGGTCG	4320
GGGACAAAGA AATGGAAGAC GCAACGGTGA ACGTCCGTCG TTATGGAAGC AAAGAAACGT	4380
CTGTCGAAGA TTTATCAATT TTCATTGACA GCATGGCTGC TGAAGTTCAC AATTACAGCC	4440
GTTAAAAAAA GAGAAaAAGT CAAAAGGACT TCTGAATTCA GAAGTCCTTT TGACTTTTTT	4500
CAACAACAAA ATAATTAAAT ATTTCTTAG AGCTGGGAAT AATCCTGATT TTTTGACTAT	4560
TTTTTGTTAC AATGAGAATG CGAAACTTTT TAGTGAAGGG ACTAGAGGTA TGAAGAAAAA	4620
CTCATTTATA GnTrAaATAA AAAAAaCAGA TGCCGCAGCA AAGAGTCAAC GAGGTAGAAA	4680
GCCACATATT CCTTTTCGAT TGAACCTATT ATTTTTGTG ATATTCACCT TATTTGTTTC	4740
ATTAATTGTC CGTTTAGGGT ATTTACAGAT TGTAAGAGGA GAAGAGTTTA ACAAAAAAAT	4800
CACTGCCAAT TCTTCACTAC AAATCACGAC ACCATCCCCT CGAGGACAAA TTTATGATTC	4860
ACAAGGAAAA GTCTTGGTTT CTAACAAAGC CAATTTAGCA ATCACATACA CACGAGGAAA	4920
AAATATTGAA GGAAAAGATA TTTTACCGAT TGCGAATAAA GTTAACGAAC TTATTAACGT	4980
TCCAGTTGAT CCAAACCTAA CGGATCGTGA CAAAAAGAT TATTGGTTAG CGAATCCTGA	5040
AAATTTAAAA GCAGCCCAAT CCCGTTTAA CAGTCAAGAT AAAGAAGATG AAAAAGGCAA	5100
CAAAATTACG GACGAAGGCA CATTATATGC CAAAGCAGTT GAAAAAGTCA CCCCTGAAGA	5160
AATTGCGTTT GATGACCGAA CCTTACAAGC AGTAACGATT TTTAAACGAA TGAATGCAGC	5220
ATCACAAATG AACACGGTCT TCATCAAGAA TGAAGGGGTA ACAGAAGGTG AAATTGCGAC	5280
AATTGGTGAA CACACAGCGG AAATATCTGG TGTTTCAACA GGGACTGATT GGACGCGTGA	5340
TTATTCTCAA AGTGGCGCAT TACGTAGCCT GTTGGGCACc GTTCAACTG AAAAACAAGG	5400
ATTgCCAGCC GAAGAaGTCG ATGanTATTt GAAAAaGGC tATGCaCGAA aTGACCGTGT	5460
CGGGACTAGC TATTTAGAAA AACAAATATGA AGACGTTCTA CAAGGTAAAA AAGCAAAATC	5520
AGAAGTCGTT TTAGACAATA ACGGAAAGAT TGTTTCACAA ACCCCAATTT CCAAAaGGGGA	5580
AAAAGGCTCC AACTTAAAT TAACGATTGA TTCAAATTC CAAAACAAAG TTGATGAAAT	5640
TTTACAACGA AACTATTAC AAATAGTCAA AACCATTGGA CCATACTCAG AAAATGCATA	5700

TGTCGTTGCT ATGAACCCTC AACAGGAGC TATTTTAGCG ATGTCCGGAT TTCATCATGA	5760
TTTAGCAACT GGAGAAGTCA CACCTAATCC TTTAGCACCT ATTTTGAATT CTGAAGTTCC	5820
TGGATCAGTT GTGAAAGCAG GTACTTTAAC AGCTGGTTAT GAAACAGGCG TCATTAAAGG	5880
CAATGACGTT TTAACAGATG AAGCTATTTT ACTAGCAGGT AGTAATCCAA AAGCTTCTTG	5940
GTGGAATGCT TCTGGCGGAA CAACCATGCA ATTAAGTGGC GAGCAAGCGT TAGAATATTC	6000
TTCGAATGCT TATATGATGA AATTAGTCTT TAAAATGATG GGTGTGAATT ATTATCCCAA	6060
TATGATTTTC CTTATGAAG TAGGAGATGA CACTGTCTTT AAAGAATTAA GaAAAGCTTT	6120
TGCAGAATAT GGTATGGGAA CGAAACTGG AATTGATATA CCAGGAGAAA CAACAGGAAT	6180
TCAArACAAG GACTTTAAGG ATTCGTCATC tGCGCCaCAA GCGGTAATC TTCTTGACCT	6240
ATCTTTTGGA CAATATGAT	6259

## (2) INFORMATION FOR SEQ ID NO: 130:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9767 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

CGCTGACTCA ATTAGCGTAA TGGAGTTTGT TTTAGAACTT GAAGATGAAT TTGGAACAGA	60
GATTTCAAGAT GAAGATGCAG AAAAAATTGA AACAGTTGGT GCCGCTGTAG ATTATATTGT	120
GAGTAATTCA TAAAAAATAA TAAGACAAAG TCGATTAACA TTCCTGTAA TCGACTTTTT	180
TTGCGTACAT AATGGTTATT TTATTAAGCA ATAAGGCTAT TTTTACACAT ATTTTGTAA	240
AAAAATCAT ACATGTAAAA AAATAACTGC GAAAAAGTT ATTTTATCTT GCTTTTATG	300
GAATAATTCT ATAAATTAT GAAAAACAA TGAGCTGTAA ATTAACATTG ATAAAAATTA	360
GAAAGTGCTA TCATTAAATA TTAGTTAATT TATATAATGT TTTTGAAAGG ACGTAAGGGG	420
TGCGTATGGA TGCAAGGAAA TGAACAATTA CTAGAGGTTT CAGGATTAGA AACAGCATTT	480
AGAATCAAAG ATGATTATTA TAATGCCGTC GATGATGTTT CATTTGAATT GAAGAAAAAT	540
GAGATTCTGG CAATCGTCGG AGAATCTGGT TGTGGAAAAA GTACCTTAGC AACAACGATT	600
ATGGGCTTAC ATGATCCCAA TCATACCAA ATTACTGGGG AAATTTTATA CAACAATTTA	660
AACTTAACTA CGTTTAATGA AACATTGTAC AACAAAGTCA GAGGAAACGA TATTGGGATG	720
ATTTTCCAAG ATCCGTTATC TGCACTGAAT CCGCTCATGC GAATTGAAGA CCAAATTAAA	780
GAAAGTCTCA CGTATCATAC CAAAATGACC GCACAACAAA AACAAGCACG GGCCATTGAA	840
CTGTTAGATC AAGTGGGGAT TCCCAACCCA GAGCGTGTCT GGAAACAGTA TCCTCaTGaA	900
TTGTCTGGcG GcATGCGCcA ACGGGTGaTT aTTGcCATTG CAATCGCGTG tAAACCACCG	960
aTTTTaATTG cmGrTGaACC AACAACGGCC TTAGACGTTA CCATTCAGGC ACAAATTCTC	1020

GATTTGTTAA AAGATTTACm AGaAGAAAcA GGGAGcGGCA TTATTTTAAat TACCCATGAC	1080
TTAGGTGTAG TTGCGGAAAT GGCTGATCGT GTAGCGGTTA TGTATGCGGG TCAGTTTGTA	1140
GAAGTAGCCA CTGCTGAAGA ATTGTTTAGT CATCCGAAAC ATCCATATAC ACGTTCCTTA	1200
CTTCAGTCCA TTCCCCAAGA GAATTCAGAT GAAGAAGAAC TTCATGTGAT TGAAGGGATT	1260
GTTCTTCGT TAACCAAATT ACCTAGAAAA GGTTGTCGGT TTGCCCCACG AATTCCTTGG	1320
ATTGATGCAT CGGAACATGA AGAAAATCCA ACATTACATG AAGTTGCACC GAATCATTAT	1380
GTTCTGTGTA CATGTTATAA ACATTTTCAT TTTAGAGACG GGGAAGGGGA AGCGTAATGA	1440
CAGAAATTAT TCAAAATAAA GATTTAAAAG TTCATTATCC AATTCGCAGC GGCTTCTTCA	1500
ATCGAATCAC TGATCATGTT TTGGCGGTTG ATGGTGTGGA TTTCATGATT GAACAAGGTA	1560
AAACGTATGG GTTAGTTGGT GAATCGGGTT CTGGAAAATC AACGACAGGA AAAGCAATTA	1620
TTGTTTATAA AAAAATTACA AATGGTGAAA TTATTTATCA AGGACAAGAT GTAACCAAGC	1680
CGCGAAGTCG AAAAGCGATT GGATACAACA AAGATGTCCA AATGATTTTC CAAGATTCGA	1740
TGTCAAGTTT AAACCCTAAA AAACGAGTAC TAGATATTAT CGCCGAACCA ATTCGAAATT	1800
TTGAACGGTT AAGTGACCAA GAAGAAAAGA AAAAAGTCAA AAGTTTGTTG GATATTGTGG	1860
GGATGCCAGA AGATGCGTTG TACAAATATC CTCATGAATT TTCTGGTGGA CAACGCCAAC	1920
GTTTAGGCGT CGCCCGAGCG GTGGCTACCA GCCCCAAATT AATTATTGCA GATGAACCTG	1980
TTTCAGCACT GGACTTATCT GTGCAAGCAC AGGTATTGAA TTTTATGAAA AACATTCAAC	2040
AGGAATATGG GTTAAGTTAC TTATTTATTT CCCATGATTT GGGCGTAGTT AAACATATGT	2100
GTGACAACAT TGCCATTATG TACAAAGGCC GTTTTGTTGA AATCGGTACA CGCCAAGATA	2160
TTTACTACTAA TCCACAACAC ATTTATACCA AACGCTTACT TTCAGCAATT CCTAAAATTG	2220
ACGTAGCGAA CCGAGAAGCT CACAAAGAAG AACGAAGAAG AGTTGAGCAA GAGTATCGGG	2280
AAAATCATAA AGACTATTAT GATGAAAATG GCGGGGTTTA TAACTTACAT GCGATTAGTC	2340
CAACCCACCA AGTAGCGTTA AAAAATGGAG GTGCTGAATA ATGTGAAAA CAATTTTACG	2400
CCGTTTGTTA TTGATGATTC CGCAAGTGAT TATTTTAAGT GTCTTGATTT TCCTGTTGGC	2460
TAAAATGATG CCTGGTGATC CGTTTACAGG ATTGATTAAT CCGAACCAAG ACCCGAAAGT	2520
GATTGAAGCC ATGCGCGAAg cTGCCGGGTT AAATGATCCT TGGTACGAAC AATATTTCCG	2580
TTGGATTGGC AATGCCTTAC ATGGAGATTT TGGGCATAGT TTTATCTTTA AATTACCTGT	2640
GTCCACGCTC ATTGCTGGTC GGGTGGGGAA TACGATTGCA TTGGCCGCTG TGTCAGTTAT	2700
AATCACGTAT TTAATCGCCA TTCCTTTCGG TTTAATTGCT GGCCGTTACC AAAATTCATG	2760
GTTTGACAAA ATGGTGGTTA TCTACAACTT CTTCAATTTT GCCGTGCCAT TATTCATCTT	2820
TGCCTTGATT ATGCTCTTTA TTTTGGGTA TCGCTTAGAC TGGTTCCCAA CGAGTGAAC	2880
CGTGACTGTT GGTTTGGCTG AGGGAACGTG GCCGTATTAT TTAGATAAAC TGAAGCATT	2940
AATTTTACCA GGGGTCACCTC AAGCGTTGCT AGGGACAGCG GTCACATTC AATATTTGCG	3000

CAGTGAAGTG	ATTGATGTTA	AAAACATGGA	CTTTGTTTCGT	ACTGCTCGTT	CAAAAGGAGT	3060
GCCAACGAAT	AAAATTTTTA	ATCGACACAT	TTTTAGAAAT	GCAGCATTAC	CAATTGCCTC	3120
TCAATTAGGT	TATGAAATTA	CTGCTTTGAT	TGCCGGATCA	GTCGTTATTG	AAAAGATTTT	3180
TGCTTTCCCA	GGAATTGGAA	AATTATTTAT	TGATAGTATC	ATTCAACGTG	ACTATTCCGT	3240
TATTACCGCT	CTCGTGTTGA	TTTTGGGATT	GGCAACCCTT	ATCGGAACCT	TAATTTCTGA	3300
TATTGTAATG	AGTATTGTTG	ATCCAAGAAT	TCGAATTCAG	TAGAAAAAAC	ACGAAGAAAG	3360
GGTGATAAAA	ATGGAATTGA	CAGAAGAAAA	AAGAGAAGAA	GTCTTACAAG	AAAGTATTCC	3420
ACCAATGGGC	TTTCGGATGA	TTGCGCGAGA	ATTCGTCAAA	GAAAAAATGG	CGATGTTTTTC	3480
TTTAATCCTT	TTGGTCATTA	TCTTATTAGC	CGTTTTTATC	GGTTCGTTAG	TGTTAGACCA	3540
AAGTGCCGTT	ATGCATGTAA	GTATTTTAGA	TAAATATGCG	GAACCAGGTA	CCGTTACAAT	3600
GAATGGCACC	AAGTTCATTT	TAGGTGCCGA	TGAAGGGGGT	CGTGATGTTT	TAGGACAGTT	3660
GATTATCGGG	GCTAGGAACT	CGATTTTAAT	CGGTTTTGCC	ATTACGATTA	TTACGTCTAT	3720
TATCGGGGTC	GGTTTAGGGA	TTATTTCCGG	TTATTATGGA	GGCATGATTG	ATAACATTTT	3780
GATGCGTATC	GTGGATTTTA	TTATGATTTT	GCCAATTATG	TTAATTATTA	TTGTGTTTCGT	3840
TTCCGTGATT	CCTAAATATA	GTATCTGGTC	CTTTATTTGG	ATTATGTGTG	CGTTTTATTG	3900
GGTTGAAAAA	GCGCGGTTGT	TCCGCAGTAA	AACGTTATCT	GAAGGGCGCC	GAGATTACGT	3960
TAGTGCCTCT	AAGACAATGG	GAACAAGCGA	TTTTAAAATT	ATGTTTCGAG	AAATTATGCC	4020
GAATTTAAGT	TCATTAATTA	TTACGAACTT	AACGATTAAT	TTCGCGGCCA	ATATCGGGAT	4080
CGAAACAACG	TTAACTTTCT	TAGGCTTCGG	GTTACCACAA	AGTGTCCCAA	GTTTAGGAAC	4140
CTTAAtTGgc	TACGCCAGCA	GTGGGGATGT	ATTAGTCAAT	AAACAATGGG	TTTGGTTACC	4200
TGCATCAATT	CTAATTTTAG	TCTTGATGTT	AAGTATAAAC	TATGTTGGAC	AAGCATTTAA	4260
GCGCTCAGCA	GATGCACGAC	AACGATTAGG	CTAATTAGCC	AAAAAATGAG	GGAGGAAAAG	4320
AGATGAACAA	GAAACGGATT	TTAGGTGCAA	TCACGTTAGC	TTCTGTGTTA	GTATTCGGGT	4380
TAGCTGCATG	TGGTGGCGGC	AATAAAGGCG	GGGGCAATAA	AGCAACGGAA	ACAGAAGACA	4440
TTTCAAAAAT	GCCAATCGCT	GTAAAAAATG	ATAAAAAAGC	AATTGATGGC	GGTACATTAG	4500
ATGTCGCTGT	AGTTATGGAT	ACACAATTCC	AAGGACTTTT	CCAGCAAGAA	TTTATCAAG	4560
ACAACTATGA	TGCACAATAC	ATGCTTCCAA	CGGTACAGCC	ATTATTTAAC	AATGATGCAG	4620
ACTTTAAGAT	TGTCGATGGG	GGTCcTGCGG	ATCTGAAATT	AGATGAAGAT	GCCAATACAG	4680
CAACCATTAA	ATTACGTGAC	AATTTGAAAT	GGTCTGACGG	TAAAGATGTG	ACAGCCGATG	4740
ACGTGATTTT	CTCTTATGAA	GTCATTGGTC	ATAAAGACTA	TACAGGGATT	CGTTATGATG	4800
ATAACTTTAC	GAATATTGTT	GGCATGGAAG	ACTACCATGA	TGGTAAATCG	CCAACCATTT	4860
CTGGCATAGA	AAAAGTCAAT	GATAAAGAAG	TTAAAATCAC	TTATAAAGAA	GTTCACCCAG	4920
GAATGCAACA	ATTAGGTGGC	GGTGTTTGGG	GCTCAGTTTT	ACCAAAACAT	GCCTTTGAAG	4980

GAATTGCTGT TAAAGACATG GAATCAAGCG ATGCAGTTCG TAAAAACCT GTGACTATTG	5040
GACCATACTA CATGAGTAAT ATTGTGACAG GTGAATCTGT TGAATACCTA CCAAATGAGC	5100
ATTACTACGG TGGTAAACCT AAATTAGATA AATTAGTGTT CAAATCTGTT CCTTCTGCGA	5160
GCATTGTAGA AGCGATGAAA GCGAAACAAT ACGATATTGC ATTATCAATG CCAACAGATA	5220
CGTATCCAAC ATACAAAGAT ACTGAAGGGT ATCAAATCTT AGGACGTCCC GAACAAGCCT	5280
ACACGTATAT TGGCTTTAAA ATGGGTACGT TTGACAAAGA AACAAATACA GTGAAATACA	5340
ATCCAAAAGC TAAAATGGCA GATAAAAGCT TACGTCAAGC CATGGGCTAT GCAATTGACA	5400
ATGATGCAGT CGGCCAAAAA TTCTACAACG GCTTACGAAC AGGGGCAACA ACGTTAATCC	5460
CACCAGTCTT CAAGAGCTTG CATGATAGCG AAgcGaAAGG CTATACGCTT GATTTAGACA	5520
AAGCGAAAAA ATTATTAGAC GATGCTGGTT ATAAAGACGT AGACGGCGAT GGCATTGCGG	5580
AAGACAAAGA AGGCAAACCA CTAGAAATCA AGTTTGCTTC AATGTCAGGC GCGGAACTG	5640
CACAACCACT TGCTGATTAC TATGTCCAAC AATGGAAAGA AATTGGCTTA AACGTAACGT	5700
ATACAACAGG ACGCTTAATT GATTTCCAAG CATTCTATGA TAAATTGAAA AATGATGACC	5760
CAGAAGTAGA TATCTATCAA GGC CGTGGG GCACAGGTT AGATCCTTCA CCAACCGGCT	5820
TATATGGTCC AAACCTCAGCC TTTAACTATA CACGTTTTGA GTCAGAAGAA AATACTAAAT	5880
TACTTGATGC GATTGATTCA AAAGCATCAT TTGATGAAGA AAAACGTAAA AAAGCCTTCT	5940
ACGATTGGCA AGAGTATGCC ATTGATGAAG CGTTTGTAAT CCCAACGCTT TACAGAAATG	6000
AAGTCTTGCC TGTCACACAC CGTGTAGTTG ACTTTACTTG GGCAGTTGAT ACGAAAGATA	6060
ATCCATGGGC AACGGTGGGT GTCACAGCAG ACTCACGGAA ATAAAAAAC AGATTCTAGA	6120
AGACATAAAG CTCTTCTAGA ATCTGTTTTT TTGGATAGGA TAGTTTCTTA GGCCATCTTA	6180
TTTTATGTGA AAATTTTTGT TACTTCAAGC AATTACAATT TATATGTATG CGTTATTTTT	6240
TTGTTATAGA TGAATTGTTT TTTATGCAGT ATGGTTTTTG ATGGGAGGTT CTGTTTTTTT	6300
ATTTTATTTA GTGTGTATAT CTACATGTTT GTTATGAAAT GTTATTTTTT TTATTATATA	6360
TATAATACTC CCTTTACGTA TAAAGAACGA TAAGTTACAA TTAGATTAAG TTTTAGTTAG	6420
TTTTTGAAGG TGTGTTTGGT ATTGTTTATA ACTAAACTT ATATTTAAAG AAAGGAGTGA	6480
ATGGATATGC CACTTTATAT GGAGAATTAT TTGCTTGCAT TGCAATCAA ATATTCACAA	6540
GAAATAACTG TAAAATGGGT TTGGGCAGTT GTAATATTAC TTTCAGTGAT TGGCTGTATT	6600
GGTTATGCAG TTTACTGTAG TTGGGTAGGT GGAACATTTG CTGGGAGTAT TAAAATTGGT	6660
GTTCCCGACT TGGTACACGT TACCTTCAAT TGTAACGTT GATATTTAAG GAGGAAGGAG	6720
CCTCTATTTT ATGGAGGCTT CTTGTTAAGA AACTTATGAT AAGAATTGAA AATTTAACAA	6780
AAAAATTTGA TAGGAAAGTT TTAGATAATA TAACTATCTC ATTGCCTAAA AACCGAGTTA	6840
GTGTGAnTGT AGGTATAAAT GGAAGTGGGA AAACACTTTT ATTAGATTGT AATGTAGGAT	6900
TGAAAGATGC AACGTCTGGA AAGGTATTTA TTGAGAGCTA TTCTAATGAT TCTGAAAAGT	6960

TTAAAGAATG TATATTCTAT ATACCATCGG AATTTTATCT ACCTAATTAT ATGACGGGTA	7020
AAGAGTACTT AAACTTTGTT TTATCAAGAT ATCGCTGTAG TGATATAGAA CGGATAGATG	7080
ATTTTTTAGA ACTATTTGAT TTAAAATTTG CTGGAACATA TTTAATAGAA TCGTACTCAT	7140
TTGGTATGAA GAAAAAATA CAAATAGTtG CAGCAGCACT AGCTAACACT GATTATATAT	7200
TAGGTGATGA AATATTTAAT GGACTTGACT TTGAGACAAC ATTGCTTACG TTAGAGCTGT	7260
TTGAAAATTT ATCTAGAGAA GTAGGGATAG TGATTATTTT TCATAATAAA TTAATTATTG	7320
AAAGGTTTTC AGAGAATATT TTATTGATGT CCAATGGCAA TTTAACTCCA TTTTtagggg	7380
CGTCAGAGAA TTTAGAAAAA GAAGTGATAA GTACGGAGAA AATTCATGAA AAAATTAAAT	7440
ACATCAAAGG ATATCATCCT ATTAATCGAG TTATTTATTG ATAATGCTTT ACGAGGTTTG	7500
CTCAAGCACT CTATTTTGTC AAAAAAGAGT ACTCGCATCG TAATAGGTGT GTGTGATTTT	7560
TTTATCTATT TTGCATACCT TTTTTTTAAT ATGAGTGAAC TAGCTAGAAT TGTACCAGAT	7620
TCAGAAAAAA TTAGTCATAT TTTGATAGAA CAAGGTCGAA AAATAACTTT TTATAGTTAT	7680
TTTAGTAATA CATTTGTTTT AGGTATAATA GCATATATTT TAGTTGaTAA TACTGTTGCa	7740
TTAGATAAAA ACTCCTTATT TTTTGTGAAA ACATTGCctT TTAGAAAGAA AGACATCAGC	7800
TTATCTTTTA TGCTATTCAA ATTTGTAATT ATGATATTGC TTTATGAATT AGTTATGATC	7860
ATTTCTACGC CAGCTATTAA ACTTGTGACG ACTGTTCTTA TTGAATATGT CATTTTTTTT	7920
ATTGTTTcAGC ATCTTTTTTA TCTAGTGGTA ATTGGGGTGA TTGAGTTTAT ACATTGTTTG	7980
TTTACATTCT TTTTGAAACG AAGAGTGAAA CAAATTAATA GTCTCAAAGT GATAGGTGAT	8040
AGTTTATTAA TGATTTTTGC TACTTTTTAT TTTTTtGATT TTCGTTATGG TTTAGAATTA	8100
TTTCTAGCAA ACCAATTGTG GAGTATCTCC TATATGATTC CTATCACTTT TTTTCTAATA	8160
GTGGTATGTT TAATGCTTAT TAGTATAGGC CTATTAAGAC TAATTAGTAT TCTTGAAAT	8220
CAAATGTCTC AACATTCAAA ATATATATAC ATTCCATTCT TAAATAAAGT tATATTTCGT	8280
TATAAAGCGA ATTGGTATTT TATTAGTTTT ACTACAGTTG TTATGTTGGT TATTTTTTTT	8340
CAAAGTGGTT TAAGAACAAT GTTATTcATT TTGACAACtG TTATGGCCTT TTCAGGCGTT	8400
TTATTATTAA GCTATGGAGA TATTACTGCT GATTTcAGAA AACAATATGA TTTATTGAGA	8460
ATAAAAATTA GAAATGAATG GTTGAGTCmA CTATTGCKAG TGATAATGTT GGCAATGCCA	8520
TtGtTACTAC TTgtTTTTGG GGATTtGGGG AGTTAGCTCA ACTAATTACA GCGTTGTcGT	8580
TATCATTAGT AGCTATAATT TTGGGATATG TATTTCCAAA ATCACAAGGA AGTTTAAATG	8640
AAACGACtTC ATTATTGTTA TTATTTATTG TATTTGTTCT AGTTAGTTTA TTGACAAATC	8700
GCTCTTTTGG TTGGTTGATT CTAATTGTTT TAGTAnTCTT GCATTTGTTA GTAATTAGA	8760
AGGTGAGAAA TGAAAAAGAA TAGTAAATAT TTAGGAAGTA TAACAGTTAT CTATTTGTTA	8820
GTTCTTGTGA TCGAAACATT ATGGGAGCTA TTCCATACCA GTCGCACAGA ATCAACGAAA	8880
GTCACTACCT TATTAGGAAT TACGATTGAT AATCGTATAA GTAAACATGA AATTTCAACT	8940

ACTTTTGGTT TGACAATTAA GGTATTAGTA CTCTATTTGT TGTATTGTT AGTTGTCTAT	9000
ATTTTGACTG GGTATTTTGG AAAAAAGAAA ATATGATTAT TAAGTTATCA ATATTGTATG	9060
TGTTGTCAAT ATATGTCGTG TTACTTGTAC ATGAGTGGTT ACATTTTATT TTAGCCmAAT	9120
TCTTCAAATA CAATGcTTAT ATAAAGAGAG TTGGTCTATA CCCATTTAAA GTGGTGTATA	9180
CGAACAGAAA TAATCCtCTA GATAATTTAT TGATTTcAGC GATATCTCCG CTTTTCTTGG	9240
TTATAGTAGG AATAATTTTG CCATTAAATT ATTATACTGT GATTTTAAAA GTATCTTGTA	9300
TTAGTAATAT TTTTAATCTC TTACCGTTTA CTGCAGATGG AGAAATAATT TTGCTTTCAA	9360
TTTTTCAAAT TTTTAGGAGG AGAAATAAAT GAAAAAAGA AAAGTGTTCa TTATTTTCT	9420
TGTTAGTTTG ATTATATTTG TTTTGGTTA TTACTTAGCG TTTAATCAAA ATAAGTATGT	9480
ATCATATAAT GACGAGACAA AACAGTTTGT ACATACTGAT ACTTCAAAAA TAATGGAAAA	9540
TCTGGTGAAC AGACAACAAG GAATTTATTA TTTTGGCTTT CCGACTTGTC CTTGGTGTCT	9600
AGAATTATTA CCAATACTAG ATAAAGAATT AGAGAAGGAA AGTATGAATG CTTATGCTGT	9660
AAATACTCGT GGTGATGACT ATACaGAAAA TGATGAkGAA TTATTACaAA AATTTkATCa	9720
AAAATATACT GGGGATGAGT CGCTTTCTGT ACCTTTTATT GTTGCTA	9767

## (2) INFORMATION FOR SEQ ID NO: 131:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7900 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

AAAATCGTTT TGGATTTTTT CTCGnTnCT TTCTATTTAA GTTATTTTAG TTGGATATCG	60
TTATTTAATA AGGCTATTAT ATGGTATAAT GATAATTAAC TGAAGTGTGG AGGAAATAAA	120
ATGTTATCAA TTGTTGCCC TTGTTATAAT GAGGAAGCAG CGATTCCGCg TTTTTTgAAG	180
AAGTAGAAAA AATTAGCCAG AAACGTGCTC ATTCAGTAGA ATATATATTT GTGAATGATG	240
GCTCAAAAAGA CAATACGTTA GCTGTTTTAA GACAGCTTTA TCGCGAACAT CCAGAGAAGG	300
TCCGTTATCT TTCTTTTTCA CGAAATTTTG GGAAAGAAGC GGGCTTATAC GCTGGTTTAA	360
AAGAAGCAAC AGGTGACTTA GTGACAGTCA TGGACGTTGA TTTACAAGAC CCGCCTGAGT	420
TATTACCTCA AATGATTGAG ATGATTGAAA CCAGTACAGA TTTAGATTGT GTTGGCACAC	480
GACGCATAAC AAGAGACGGC GAACCGCCCA TTCGTAGCTT TTTTGACGG ATGTTCTATA	540
AACTCATTAA TCGAATTGCT GAGACGGAAA TGGTGGATGG TGCCCGTGAC TTTCGTGTCA	600
TGACCCGGCA AATGGTCGAT GCGATTTTAG AGTTATCTGA GTATAATCGT TTCTCAAAG	660
GGATTTTTAG TTGGGTCGGC TTTAAACAG AATATATTGA ATTTAAAAAT CGAGAACGAA	720
TTGCTGGAGA AACCTCTTGG TCTTCTGGA GTCTATTAAG TTATTCAATT GACGGGATTG	780
TCAATTTTTT AGAAACACCA TTGAATATTG CTTCATACGT TGGCGCCTT TCGTGTATTG	840



GTTCTGCACT	AGCAATGTTA	GTGATTATCT	TTCGAACATT	AGTTAATGGT	GATCCAACGA	900
GCGGCTGGCC	TTCAATGGTT	TGTATTGTAC	TTTTTGTTAGG	TGGTTTGCAA	CTATTATGTT	960
TAGGAATTAT	TGGCAAATAT	ATCGGTAAAA	TTTTCTTAGA	AACGAAAAAA	CGACCAATTT	1020
ATATTGTAAA	GGAATCAGAA	AAAGATTCTA	AAAAATAACC	AAAAAAGCAG	AGAAAAATTC	1080
TCTGTTTTTT	TGTTGTGCAT	TAAAAAAGT	GTGAATAGAC	TTTTCAGAAA	AACAACATTT	1140
TTTACATAAA	AAATACAATA	ACTTTACCTA	TTTTTATCTT	AATGAAACAG	CTTTGCTCTC	1200
GCGGTTGCAA	GAAATAAATG	TTAAATCAAA	AAGGTATGCT	ATTAAATGTT	TTTTTGGAGG	1260
TTGGGTcATG	ATAGAAGTTA	TCGACTTGAA	AAAAGTATTC	GATAATGGGT	TTGAAGCGTT	1320
AAAATCGGTC	AATTTTACGA	TTGAGCAAGG	GGAAGTGGTT	TGTTTGTTAG	GACCAAGTGG	1380
GTGTGGGAAA	TCCACAATTT	TAACTTAAT	TGCAGGTTTA	TTGCATCCAA	GTGATGGAGA	1440
TATTCAGTTT	CGACAGCAAT	CGGTTGTCAA	AACAGCACCA	AAAGATCGGA	ACATCGGGTT	1500
CGTTTTTCAA	AACTATGCCT	TATATCCACA	TATGACTGTT	CTAGAAAACG	TAATGTTTCC	1560
TTTGACGGTA	GGAAGTAAGA	AAGTTCCCAA	GGCAGAGGCA	CAAGCTATTG	CCGAAGAATA	1620
TATGAACTA	ACGAATATTG	AAGAGTTAAG	CCATAAAAAG	CCAGGCACAC	TTTCTGGTGG	1680
TCAACAACAG	CGTGTGCGA	TTACCCGGGC	GTTAGTTCAA	AAACCAGATG	TTTTATTATT	1740
AGATGAGCCT	TTAAGTAACT	TGGATGCACG	TTTGCGTTTG	AAAATTCGTG	AAGAAATCCG	1800
CCGCTTGGTG	AAAGAAGTAG	GGATTACAAC	AATCTTTGTA	ACGCACGACC	AAGAGGAAGC	1860
GCTATCAATC	AGTGATAAAA	TTATTTTATT	AAACGAAGGG	GTTATTCAAC	AAAATGATGA	1920
ACCTCAAAAC	CTTTACTTAG	AACCAAATAA	CTTATTTGTT	GCTCAGTTCA	TAGGTAACCC	1980
AATTATTAAT	TTATTGTCTG	TTGAAGTGAA	AGACGGCAAA	ATGTACCACA	AAAGTTTGA	2040
AATTCCGCTT	GAGCGTTTTG	AGCAAGCACG	CTTTAAAATG	CCAATGACTG	ATGGGAAATA	2100
TACGTTTGCT	TCTCGTCCgG	AAGATGTGGT	ACCAGCTGAA	ACAGGTCTCT	TTACCACAAC	2160
GACAGATTTA	GTGGAATTGA	TTGGTCGAGA	ACGTATTTTA	CGATTTACAT	TAGGAAATGA	2220
ACAAGTGAAA	TCAATTGTAA	GTGTAGAAGA	AGCGATTGAA	GAAGGAGACA	CTTTATCTTT	2280
TGATTTTTCA	TATAAAAAAG	TATTTATCTT	TAACGAAGCG	GGAGACCGGG	TTTACTAATG	2340
AAAAAATACA	ATCCAGAAAA	TCAACCAAAA	GCATGGCTCT	TCCTTCTGCC	ATCATTAGGA	2400
ATCATTTTAC	TGTTTAGCGT	TTATCCACTG	TTTCGTTCTT	TGTGGATGAG	TTTCCAAAAA	2460
GGTTCCTTAA	TTAACCAACG	cTACGCTGGT	TTAGAAAATT	ATCAACGTGT	ATTGAATGAT	2520
CCAATCTTCT	ATAAAGCGTT	AAAAAATACA	GCGCTCTATG	CATTTGCTGT	GGTACCGATT	2580
GCCTTGATAA	TTTCCCTAGC	AATTGCTTGG	ATTATCTTTG	AAAAAGTTAA	ACATAAGAGT	2640
TTCTTTGAAA	CGATTTTCTT	TATGCCTTAT	GTAACGAGTA	CGATTGCCAT	TGGGATTGTC	2700
TTCCGTTACT	TCTTTAATGG	CGATTATGGA	ATTGTCAATT	ACGTCTTAGG	CTTTTTCGGC	2760
ATTCCTTCTG	TCAACTGGCT	AGATAATGTT	CAAATGAGTA	TGCCAACATT	GATTATTTTT	2820

GGGGTTTGGG	CGAGTTTAGC	ATTTAATATT	ATTATTTTGT	TGGCTGGGTT	GAGAAATATT	2880
GATGAAGAAC	ATTTTAAAAT	TGCAAAAATG	TTTGGCGCCT	CAGACGGCGA	AATTTTCCGA	2940
CGTATTACGT	TTCCGCAGCT	AGTCCCGACC	ATTGCCTTTT	TATTAACGGT	CAACTTAATT	3000
GGCGCCTTTA	AAGTTTATAC	ACAAGTTTAT	GCTTTATTCT	GTGGCCGTGC	GGGGATTGCC	3060
AATAGTGCCA	CAACCGCAGT	GTACTATATT	TATGACAAAT	TCCACATTGC	AGGACGTCCC	3120
GGAATTGCGA	TGGCCGCCAC	AGTGATTTTA	TTTGTGATTA	TTTTAGTGGT	CACTTTCTTA	3180
CAAAATAAAC	TCTTAAAGAA	AGTGGGGCAA	TAAGCCAATG	AAAAAAGTTT	TAACGATTAT	3240
CGCATTTGTT	TTCTTAGGCA	TTTAGCCGT	TATTACGTTA	TTTCCATTG	TCTATATGAT	3300
TTTAGCTGGT	TTAATGAGTT	ACTCTGAAGC	GACCAGCATG	CCGCCAACCA	TGTTTCCGAA	3360
ACAACCACAA	TGGCAAAACT	ACACAGAGGT	CTTTCAAAA	GCACCCCTTC	TCGGTATTT	3420
CTTGAACACA	GTCTTTGTTT	CAGGCGTTAC	AACGATTGCT	ACGGTAGTCA	CCGCCGTGTT	3480
AGCTTCATTC	GCATTAACGA	GTTTGAAGTT	CCGTTTTAAA	AATGTGGTCA	TTGCTTTGAT	3540
GATTTTCGTT	TTAATGGTTC	CGTATGAATC	TATTATTTTT	ACCAATTATC	AAACGATTGC	3600
GCAATTAGGT	TTGTAAATA	CTTACAGCGC	CTTAATTATT	CCATTTTTAA	CTAGTATTTT	3660
TTATATTTAC	TATTTAAACG	GTTATTTAAA	AGGTATTCCT	GATACTTTTT	ACAAGGCTGC	3720
CAAAATTGAT	GGCGCTAGTG	ATTTAGAATA	TATTTGGCGT	ATTTTAGTAC	CAATGTCAAA	3780
ACCAGCGTTG	GTAACAGTAG	GGATTTTAAC	ATTTATTTCT	AGTTGGAATT	CTTCTTGTG	3840
GCCATTGTTG	GTGACGAACG	AGAAAAATA	CCGTCTATTG	AATAATGGAC	TATCGGCCTT	3900
TGCGACAGAG	AGCGGTAGTG	ACGTACATTT	ACAAATGGCT	GCTGCAACAT	TAACCGTTAT	3960
TCCAATTTTA	ATTATTTACT	TGATTTTCAG	AAAAGAAATT	ATCAGAGGAG	TTGCAAAAAA	4020
TGGAATCAAA	GGCTAAAACA	ACTGTGACTT	TTCACAGTGG	CATTTTAACA	ATTGGAGGAA	4080
CCGTGATTGA	AGTGGCGTAT	AAAGATGCCC	ACATCTTCTT	TGATTTTGGT	ACAGAATTTT	4140
GACCAGAATT	AGATTTGCCT	GATGATCACA	TTGAAACCTT	AATTAATAAT	CGTTTAGTGC	4200
CAGAATTAAA	AGACTTATAT	GATCCACGTT	TAGGTTATGA	ATATCATGGT	GCAGAAGACA	4260
AAGACTATCA	ACACACCGCT	GTTTTCTTAT	CTCATGCCCA	TTAGATCAT	TCACGCATGA	4320
TTAATTATTT	AGACCCAGCT	GTTCCGTTGT	ACACCCTAAA	AGAAACGAAA	ATGATTCTAA	4380
ATAGTTTAAA	TCGAAAAGGT	GATTTTTTGA	TTCCATCACC	GTTTGAAGAA	AAGAATTTTA	4440
CCCGAGAAAT	GATTGGTCTT	AATAAAAACG	ATGTGATTAA	AGTTGGCGAA	ATTCAGTGG	4500
AAATCGTGCC	TGTTGATCAT	GATGCGTACG	GTGCTTCGGC	ATTACTGATT	CGCACGCCTG	4560
ACCATTTTCAT	CACATACACA	GGTGATTTGC	GCTTACATGG	CCATAATCGT	GAAGAGACTT	4620
TAGCTTTTTG	TGAAAAAGCC	AAACATACTG	AATTATTAAT	GATGGAAGGC	GTAAGCATT	4680
GCTTCCCAGA	ACGTGAACCA	GATCCAGCCC	AAATAGCGGT	TGTCAGTGAA	GAAGATCTTG	4740
TTCAGCACTT	GGTTCGTTTA	GAAC TAGAAA	ACCCGAATCG	ACAAATTACC	TTAATGGTT	4800

ATCCAGCAAA	CGTGGaACGT	TTTGCTAAGA	TTATTGAAAA	GTCACCACGT	ACAGTCGTTT	4860
TAGAAGCAAA	TATGGcTGCG	TTGTTACTTG	AAGTATTTGG	AATAGAAGTT	CGTTATTATT	4920
ATGCTGAATC	TGGTAAAATA	CCAGAATTGA	ATCCAGCGTT	AGAAATCCCg	TATGACACGT	4980
TACTAAAGGA	CAAAACAGAC	TATTTGTGGc	AAGTTGTGAA	CCAGTTTGAC	AACCTCCAAG	5040
AAGGTAGTTT	ATACATTCAC	AGTGATGCAC	AACCGTTAGG	GGACTTTGAT	CCACAGTATC	5100
GCGTGTTTTT	AGATTTGTTG	GCTAAAAAAG	ACATTACTTT	TGTCCGCTTA	GCTTGTTTCTAG	5160
GACATGCAAT	TCCAGAAGAT	CTGGATAAAA	TTATTGCATT	GATTGAACCT	CAAGTATTGG	5220
TTCCAATCCA	TACGTTAAAA	CCAGAAAAAC	TGGAAAAACC	GTATGGTGAA	AGAATATTGC	5280
CAGAACGTGG	CGAGCAAATT	GTTTTATAAA	TTTTTTTAAG	GGAGAGAAAA	AAATGAAGTT	5340
CAAAACTCTA	GCAACAACAG	TGTTAGCAAC	CGCAGCTATT	TTCGCATTGG	GGGCTTGTGG	5400
TAACGGTAAT	GGGGCCAAAG	AATCAAACGA	TATTGTGAAA	GAAGTGAAGG	AAGATACGAC	5460
AATCACTTTC	TGGCATGCAA	TGAATGGGGT	TCAAGAAGAA	GCGTTAACAA	AATTAACGAA	5520
AGACTTCATG	AAAGAAAATC	CAAAAATTAA	AGTGAATTA	CAAAATCAAT	CTGCTTACCC	5580
TGATTTACAA	GCCAAAATCA	ATTCGACTTT	AACTTCACCA	AAAGATTTAC	CAACAATTAC	5640
GCAAGCGTAC	CCAGGCTGGT	TATGGAATGC	TGCACAAGAT	GAAATGTTAG	TGGACTTAAA	5700
ACCATATATG	GATGATGACA	CAATCGGCTG	GAAAGATGCA	GAGCCAATTC	GTGAAGTATT	5760
GTTAGACGGC	GCCAAAATCG	ACGGCAAACA	ATACGGCATT	CCATTTAATA	AATCGACAGA	5820
AATGTTATTC	TATAATGCTG	ATTTGTTGAA	AGAATATGGT	GTTGAAGTAC	CGAAAACATT	5880
AGAGGAATTA	AAAGAAGCTT	CTAAAACAAT	TTACGAAAAA	TCCAACAAAG	AAGTCGTTGG	5940
TGCTGGTTTT	GACTCGTTAA	ATAACTATTA	CGCAATTGGA	ATGAAAAACA	AAGGCGTTGA	6000
TTTTAATAAA	GACTTAGATT	TAACAAGCAA	AGATTCACAA	GAAGTCGTGG	ACTATTACCG	6060
TGATGGTATC	GAAGCAGGTT	ACTTCCGCAC	AGCTGGTTCA	GATAAATATT	TATCTGGCCC	6120
ATTTGCAAAC	AAAAAGGTAG	CAATGTTTGT	CGGTAGTATT	GCTGGTGCTG	GTTTTGTTCA	6180
AAAAGATGCT	GAAGCTGGTG	GCTATGAATA	CGGTGTTGCA	CCACGTCCTG	AAAAAATCAA	6240
CTTACAACAA	GGAACAGATA	TTTATATGTT	CGATAGTGcT	ACGCCAGAAC	AACGGACAGC	6300
GGCATTTGAA	TTCATGAAAT	TCTTAGCTAC	TCCTGATTCA	CAATTGTACT	GGGCACAACA	6360
AACAGGTTAT	ATGcCAATTT	TAGAATCTGT	TTACACAGT	GATGAGTACA	AAAATTCTAA	6420
GACAACCAAA	GTACCTGCAC	AACTTGAAAA	CGCAGTAAAA	GATTTATTCG	CTATCCCAGT	6480
AGAAGAAAAT	GCTGATTCAG	CCTATAATGA	AATGCGGACA	ATTATGGAAA	GTATTTTTCG	6540
TTTCATCAAT	AAAGACACGA	GAAAAATTATT	GAAAGATGCA	ACATCACAAT	TTGAACAAGC	6600
ATGGAACCAA	TAATCAATAA	ATAAACAGAC	GTGGGAGGTT	GGGATCTGAG	TCGTTATGGC	6660
TCGTGTTCTA	ACCTTCCTGT	TCTTTTGGGC	GGAAGAAGGA	GAGAAACAAA	ATGCAAATAA	6720
AAATTTTAGC	AACAAGTGAT	ATGCATGGGT	ACATTATGCC	AACAAGTTAT	AGTGAAAAAA	6780

AGATGGATTT ACCTTTTGGG ACCGCAAAAG CAGCAACCAT GCTGAAAAAG TTACGGGCGT	6840
CTGCCAAAGG ACCAGTTTTT CAAATTGAAA ATGGGGATTT TATTCAAGGT TCGCCGCTTA	6900
GTTACTACGT AAGAAAAGCA GAAACACATT CAGTGGCTGC CATCACTAAA ATCATTAAATC	6960
AAATGAATTA TGATGTTAGC ATTTTAGGTA ACCATGAATT TAATTACGGT TTAGATTATT	7020
TAAAGGAAAC GATTGCCAGT TATCAACAGC CAGTTTTAGC TGCCAATATT CTTGGTAAAG	7080
ATGGACAACC TTATTTCCGT CAGCCGTATG TCATCATAGA AAAACAAGGG GTTAAAGTAG	7140
CCATCCTTGG CGTGACGACT CAATATATTC CTCATTGGGA ACAGCCGGCA ACAGTCAAAG	7200
ATTTAACGTT TAAAGTGTC GTTGAAACAG CGGCGGAATA CGTGCCAAAA CTGCGGGAAG	7260
AAGCGGATCT AGTTGTTGTC GCTTATCATG GCGGTTTTGA AAAAGATTTA GAAACAGGGG	7320
AACCCACGGA ATTACTAACA GGAGAAAATG AAGGCTACGA TTTATTGGAA AAAGTTCCTG	7380
GGATTGATGC CTTAGTCACT GGTCAACCAAC ATCGAGAAAT TGCTACAAAA CTGAATGGAA	7440
TACCTGTGAT TCAACCAGGT TTTGAGGAG CTTTCGTTGG AGAAATTACT TTAGAAATTG	7500
AACCTATGGC TAAAGGCTAT CACGTTATTG GAAGTGACGC AGCTATTCAT CCTGTTGGAA	7560
ACGAACAACC AGATACAGAA GTTTTAGCTT TGACAACAGC GTTACATGAC GAGGTAGAAG	7620
AGTGGcTTGA TCAACCAAGT GGAAATGTGG AAGGGGATAT GACGATTCAA AATCCGAATG	7680
CCGTACGCCT CAAAGAACAT CCATATATAG aATTTATTAA TAACGTCCAA ATGGCTTCAA	7740
GCGGAACGGA TATTTAGGG ACTGCACTTT TTAACAATGA AGGCAAAGGG TTTAATAACC	7800
AAATTACGAT GCGCGATATC ATTACCAACT ACATTTATCC AAATACCTTG GCTGTATTGC	7860
GTGTCACTGG ACAAGACCTC CGTGAAGCGT TAGAACAAAG	7900

## (2) INFORMATION FOR SEQ ID NO: 132:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5480 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

TCCTTACATT TCATGTTATA AAATGTAACA TAGAAATTAA GAACAAAAGA TTAAGATCAG	60
TTGAAGTCCG AACAGCAAA AGTGGCGAAA ATATGTTCCG TATGTTAAAC TAATAAGTGG	120
ACATTCTGCC AGAATGGGCC GATAATTTAA GAGAAATTAG ACACTTTGAG GGGGAGTTAT	180
AAATGATCGA AAGAGAACT TCCAATACCT TTCATTTGGC TTCTAAATAT GAGCCTGCCG	240
GAGATCAGCC AGCCGCTATT GCTGAACTAG TTGACGGTGT CAAAGGTGGA GAAAAAGCTC	300
AAATTTTACT TGGAGCGACC GGAACAGGTA AAACATTAC AATCTCTAAT GTGATTCAAG	360
AAGTGAACAA ACCAACATTA GTGATTGCAC ATAATAAAAC ACTAGCTGGA CAATTATACG	420
GGGAATTTAA AGAATTTTTC CCAGATAATG CGGTCGAGTA TTTTGTGAGT TACTATGATT	480

ATTATCAACC AGAAGCATAT GTACCCTCAA GTGATACATA TATTGAAAAA GATTCCACTA	540
TTAATGATGA AATTGATAAA CTACGCCATT CAGCGACAAG TTCTTTATTG GAACGAAACG	600
ATGTCATCGT GGTAGCATCA GTTTCTTGTA TTTTGGTTT AGGGGATCCG CGGGAATACA	660
GTCAACAAGT TGTTTCATTA CGTGTGGGAA TGGAAATGGA TCGTAATGAG TTAATAAAAA	720
GCTTAGTCGA TATTCAATTC GAACGAAATG ATATTGATTT TCAACGGGGG CGTTTTCGTG	780
TCCGCGGTGA TGTGGTTGAG ATATTCCCAG CTTACGGGA TGAGCATGCG CTACGTGTTG	840
AATTTTTCGG AGACGAAATT GATCGAATTC GTGAAGTAGA TCGGTAACT GGTGAAATTG	900
TCGGCGAAAC AGAGCATGTC GCTATTTTTT CGGCAACACA CTTCGTGACG AACGAAGAAC	960
ATATGGAACA CGCCATATCG CAGATCCAAG AAGAATTAGA GGCACGTTTG AAAGTGCTTC	1020
GCAGTGAAAA TAAATTATTG GAAGCTCAAC GTTTGGAACA GCGAACCAAC TATGATATTG	1080
AGATGATGAG AGAGATGGGC TATACGTCAG GGATTGAGAA TTACTCTCGA CATATGGATG	1140
GTCGCCAAGA AGGAGAGCCG CCATATACGC TGTTGGACTT TTTCCAGAT GACTTCTTAC	1200
TTGTAATCGA TGAGTCTCAT GTAACAATGC CGCAAATTAG AGGAATGTAT AATGGCGACC	1260
GAGCACGGAA ACAAATGTTA GTGGACTATG GTTTCCGGTT ACCAAGTGCA CTTGATAACC	1320
GACCGCTTCG TTTAGAAGAA TTTGAACAAC ATGTAAACCA AATTGTTTAT GTATCGGCAA	1380
CACCAGGCCC TTATGAAATG GAACAAACAG AGACTGTTGT TCAACAAATC ATTCGACCAA	1440
CAGGTTTACT AGATCCAGAA GTAGAAATCC GCCCAATTAT GGGTCAAATt GATGACTTGG	1500
TGGGCGAAAT TCATGAaCGG ATCGAAAAAG ATCAGCGGGT GTTTGTCACA ACCTTAACGA	1560
AGAAAATGGC CGAAGATTTA ACCGACTATT TCAAAGAATT AGGCTTAAAA GTTAAATATT	1620
TGCATAGTGA TATAAAAACA CTGGAAAGAA CGGAGATTAT TCGCGATTTA CGATTAGGCG	1680
AATTTGATAT TCTAATAGGT ATTAACTTAT TACGTGAAGG AATCGATGTG CCAGAAGTTT	1740
CCTTAATTGC GATTCTAGAT GCGGATAAAG AAGGCTTCTT ACGTAGTGAA CGCTCGTTAG	1800
TGCAAACGAT GGGACGGGCT GCCCGAACG CTGAAGGAAA AGTCATTATG TATGCAGATA	1860
AAATCACCGA TTCGATGCAA CGAGCTATGG ATGAAACGGC ACGACGTCGG GCAATCCAAG	1920
AAGCATATAA TGAAGAACAT GGTATTGAAC CAAAAACAAT TATTAAAGAA ATTCGTGATT	1980
TGATTTCTAT TTCCAAAACA GCTGATAAAG ATGAAACAGT GGTTC AATTG GATAAATCAT	2040
ATAAAGATTT ATCAAGACAA GAAAAAGCTG ATTTATTAAT GAAACTTGAA AGAGAAATGA	2100
AGGATGCGGC CAAAGCATTG GACTTCGAAA CTGCAGCCAC ATTACGAGAT ACCATTCTTG	2160
AATTGAAAGC TGCCAAATAA GTCAACCAAC GGATTAAGTA AAAGATACGA TGAACCATCG	2220
TATCTTTTAC TAATCTTTTC TCTATAGAAT GACTTGAAAT GATTTGACGT AAGAGATTTA	2280
GTAAAAGGAG AACATATATG GCAATGATA AAATTGTGAT TCATGGTGCA CGCGCCCAT	2340
ACTTAAAAAA TATTGATGTC ACGATTCCCTC GTGACAAAAT GGTGTGCGTA ACTGGACTAT	2400
CTGGTTCTGG CAAAAGTTCA TTAGCGTTTG ATACGCTTTA CGCAGAAGGG CAACGACGTT	2460

ATGTAGAGAG TCTCTCTGCT TATGCGCGAC AATTTTGGG ACAGATGGAT AAACCAGATG	2520
TTGACAGCAT CGATGGTTTA AGTCCAGCTA TTTCAATTGA TCAAAAGACG ACAAGTAAAA	2580
ACCCACGCTC AACGGTGGGG ACGGTCACAG AAATCAATGA TTATTACGA TTGTTATTTG	2640
CTCGGGTAGG TCATCCAATT TGTCTAACG ATCATATTGA AATTACCAGT CAGTCTGTTG	2700
AACAGATGGT TGATAAAGTC TTAGAGCTAC CAGAGCGAAC AAAAATTCAA ATTTTGCCCC	2760
CTGTTGTGGT CAAGAAAAAA GGCCAACATA AAAAAGTGTT TGAAATGATC CAACGTGAAG	2820
GGTATGTCAG AATGCGGGTG GATGGTGAAA CCTATGATGT TAGTGAGGCA CCAGAACTTG	2880
AAAAAATAA AAAACATGAT ATTGCGATTG TGATTGACCG TATCGTTGTG AAAGAAGGCA	2940
TTGCTCTCG CTGTTTGAT TCGTTGAAG CGGCTTGCG TTTAGCTGAA GGTACGCAA	3000
TTGTTGATGT GATTGGTCAA GAAGAAATGT TGTTTAGTGA ACATTATGCT TGTCTTATT	3060
GTGGGTTTAC AGTCGGTGAA TTGGAACCGC GCTTGTTTTT ATTTAATGCA CCATTTGGCG	3120
CTTGTCCTGA TTGTGATGGT TTAGGTGTTA AATTAGAAGT GGACAAAGAC TTAGTGATTC	3180
CAGATCCAC TAAAACCTTG AGAGAAGGGG CCATTGTTCC GTGGAACCTT ATTAGTTCCC	3240
AATATTATCC ACAAATGTTG GAGCAAGCAG CCACCAGTTT TGGGATTGAT ATGGATACGC	3300
CGTTTGAAGA ATTACCTGCA GACCAACAAG AAATTATTTT AAATGGTTCT GCGGAGAAAA	3360
ACTTTCATTT TCATTATGAA AATGACTTTG GTGGTGTTCTG TGATGTGGAA GTGCCATTTG	3420
AAGGGATTTT AAAAAATATT AAACGACGTT ATCATGAAAC GAATAGCGAT TTTACACGAG	3480
ATCAAATGCG GTTATACATG ACAGAATTGA CTTGTGAAG CTGTCAAGGG TATCGTTTGA	3540
ATCCGCAAGC TTTAGCTGTA AAAATCAATG GCACGCACAT TGGTGAAGTC AGTGAATTAG	3600
CAATAAAAAA TGCGGTCCAA TTTTGAAG GTGTGTCTTT ATCTGAACAA GAAACAACGA	3660
TTGCTCGGCC AATTTTAAAA GAAGTCGAAG ATCGGTTAAC CTTCTTAAAA AATGTTGGGT	3720
TAGATTACTT AACCTTAAGT CGAGCTGCTG GTACACTTTC TGGCGGAGAA GCACAAAGAA	3780
TTGCTTAGC AACACAAATT GGATCAAAT TATCAGGCGT TCTTTACATT TTAGATGAAC	3840
CATCAATTGG TTTGCATCAA CGAGATAACG ACCGCTTGAT TGAATCATG AAAAAATGC	3900
GGGACTTAGG CAATACATTA ATTGTGGTGG AACATGATGA AGATACAATG ATGGCTTCGG	3960
ATTACTTGAT TGATGTGGGA CCTGGCGCTG GACATCTGGG CGGTGAAATT GTAGCCGCTG	4020
GTACGCCGGA AGAGGTTGCT AAAAAATCCGC ATTCATTGAC CGGGCAATAT CTTTCTGGGA	4080
AAAAAGTGAT TCCTGTACCA AAAGAACGTC GCAAAGGTAA TGGCAAAGCG ATCAAAGTCA	4140
CTGGTGCCAG TGAAAATAAT TTAATAATG TCAGTGTGA ATTTCCGCTG GGTGAATTTG	4200
TTGCTGTCAC AGGCGTTTCA GGTTCAGGGA AAAGTACTTT AGTCAATCAA ATTTTAAAAA	4260
AAGCGCTGGC CCAAAAATTA AATCGTAATT CCAATAAACC CGGGAAACAC AAAAGCATCA	4320
CTGTTATGA AGCATTGAA AAGATTGTTG ACATTGACCA AAGTCCAATT GGTGGAACAC	4380
CAAGAAGTAA TCCAGCGACT TATACTAGTG TTTTGTATGA CATTGCGGAT TTGTTTGCCC	4440

799

AAACAAATGA AGCCAAGGTT CGTGGCTATA AAAAAGTGG CTTTAGTTTT AACGTCAAAG	4500
GTGGCCCTTG TGAAGCTTGT CGCGGGGATG STATCCTCAA AATTGAAATG CACTTTTTAC	4560
CTGAGGCTTA TGTGCCTTGC GAAGTTTGTC ATGGCAACG TTATAATTCT GAGACGTTAG	4620
AAGTTCATTA TAAAGGCAAA AACATTTCTG ATATTTTAGA TATGACGGTG GAAGATGCCG	4680
TAGATTTTTT CAAACATATT CCTAAATTC ATCCTAATT GCAAACAATT GTGGATGTAG	4740
GCTTAGGCTA TGTGACATTG GGGCAGCCGG CAACTACTTT ATCTGGTGGG GAAGCACAGC	4800
GGATGAAGCT AGCGAGTGAA CTTCAATAAA ATTCTATGG GAAAACTTT TATATTCTGG	4860
ATGAACCAAC AACAGGCTTG CATACCGATG ATATTGCGCG ATTGCTTCTA GTATTAGAGC	4920
GTTTAGTGGG AGCGGGCAAT ACCGTTTTAG TGATTEAACA TAACTGGAT GTCATCAAGT	4980
CTGCGGATCA TGTGATTGAT CTAGGCCAG AAGGTGCGA CGGTGGTGGG ACGATCGTGG	5040
CGACCGGAAC ACCAGAAGAA GTTGCTAAAG TCAAAGACAG TTACACAGGA TATTATTTAA	5100
AACGTGTCTT GAAATAAACA TTAATGATTA GATAAAGAAC CTTGTTTTTA TGGGAAACAA	5160
GGTTTTCTAT TATAAATAA CAATATTTT TTAATAAAA TAAAAATAA ATGTTGACTT	5220
AATGTGTAAT TAAAGATAT ACTATGTTG TTAACAAAT AACAATGAAT ACGAAGGGAA	5280
GTCTTATATG AAAAAAACA AATAACAAC AAGTTTATTA ACAGCAGGAA TGGTCTTAGC	5340
TATCGGATC ACACCTTTTG TACATAGTAA CACATGGAAG CAGCGACCTT AATTCACAA	5400
TCAGACCATG AAGTAATG AGTTTACCA GAAAGCTAA AAGAAGCACG GGCATCAATT	5460
GGGCCGCGC AGTCCAGCGT	5480

(2) INFORMATION FOR SEQ ID NO: 133:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10391 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

CAGGCTTTTT AATGACGAAT GTCGTTCAAC AATCCATGCG CGATTGCGT CGCGATATTG	60
AAGAAAAAT CATCGTTTG CCAGTTTCTT ATTTTGATAA GAACCAACAA GGAAATATTT	120
TGTCACGGGT GACGAACGAT GTGGATGCTG TCAGCAATGC GATGCAACAA AGTTTTATCA	180
ATATTGTTTC AGCAGTCTTA GGTATTGTGA TGGCGTAGT GATGATGTTT TTAATCAATC	240
CGCTGATGGC GATTTTTTCA GTGATTATGA TTCCGTTGTC TCTGATTATT TCCAGAACAA	300
TTGTTAAAT CTCCAGAAA TATTCCAAG GAATGCAAAA TTCTTTAGGA GACTTAAATG	360
GTTATGTCCA AGAAATATG ACTGGGTTC GTGTCTTAAA ACTATATGGT CGGGAAAAAG	420
AAACCCCTGA AGGCTTTAAA CAAGTCAATC ATCGTTTAAA TGGTTTTGGT TTCAAAGCAT	480
CCTTTATCTC AGGATTAATG TTGCCATTGG TTCAGATGAC CGCTTATGGG ACCTATATCG	540
GGGTAGCTGT CTTGGTAGT TACTATGTTG TTGCTGGTGT GATCGTAGTG GGGCAATTAC	600

AAGCGTTTAT TCAATATATT TGGCAAATTA GCCAACCAAT GGGGAATATT ACGCAGTTGT	660
CTGCAGCTTT ACAAAGCGCT TCAGCTTCGA CCATGCGGAT TTTTGAAATC CTAGATGAAC	720
CAGAAGAAGA ACTTAACGAA CAAGATGTTT CTTTGCCAGA ACCTATTTTA GGCTCTGTTG	780
AATTTGAAAA TGTCAGCTTT AGTTATGACC CAGAAAAACC GTTAATTCGT AATTTGAACT	840
TTAAAGTTGA TGCGGGCCAA ATGGTTGCGA TTGTGGGACC AACTGGCGCT GGGAAAAACA	900
CCTTAATCAA CTTACTGATG CGTTTTTATG ATGTAACAGA AGGCGCCATT AAAATTGATG	960
GCATTGACAC GAAAAAATG AACCGTAGTG ATGTCCGATC TGTATTTGGA ATGGTATTGC	1020
AAGATGCTTG GTTGATATAA GGTACCATTG CAGATAACAT TCGTTTTGGG aAGTTAGATG	1080
CCACGGaTTA TGaAGTTGTC GATGCaGCGA AAACGGcCAa TGTGGATCAC TTCaTTCGGA	1140
CAATGCCAGA CGGGTATGaA ATGGAaATCa ATTCTGAGGG AGATAACGTT TCCCTTGGtC	1200
aAAAACAATT GTTGACCATT GCCCgAGCGG TAATTTCTGA TCCGAAAATT TTGATTTTAG	1260
ATGAGGCGAC TAGTTCAGTC GATACACGCT TGGAAGCCTT AATTCAAAAA GCAATGGATC	1320
GTGTTATGGA AGGACGAACG AGTTTCGTTA TTGCCCACCG CCTATCAACT ATTCGTGAAG	1380
CTGATTTAAT TCTTGTTATG AAACAAGGAG AAATCATTGA AAAAGGTACG CATCATGAGT	1440
TGCTGGAACA AGGTGGCTTC TATGAAAAAC TATACAATAG TCAATTTGCT GAAGAAGGCG	1500
ACTATGAGGA ATAAAAAGAG GTCGGGACAG AAGCGTTTAA CTCCGAGAAA TAAGAAGGAA	1560
TTTCCGAAAA TTGTTCTGTA ATTTTGGAG AATTCCGGCT TATTTCCGAA GGAGTTGCTT	1620
TTGTTCCCGC CGTTTATCAG TTTTGGAGCG TGGAGCAAAA ATCCAAAGTG ATTTTGTCC	1680
CACGCTCTTT TCTTGCTTT TTTACTATGA AAGTGAGGAA TTCTTGCTA TACTTAATGA	1740
AAAGAGGTGA CAGAGTTTGA AAAAAGAATA TTTTATTGCT GCTAGTGCAG ATGTCTATGG	1800
AAACGTACAA CTAGCAAAGG ATAGTAATAT TTGGTTTCAG TCGGTCTTAA GAGGTGACAG	1860
TAACACAATT ACCATTGGTG AAGGTAGTAA TATTCAAGAT GGCACGATTA TTCATGTTGA	1920
TGAAGATGCC CCGACAATTG TAGGCAAGTA TGTAACGGTG GGCCATCAAT GCATGTTACA	1980
CGGTTGTAAA ATCGGTGATG GTGCTTTGAT TGGCATGGGT TCAACCATTG TAAATAATGC	2040
TATTATTGGA GAAAACAGCC TGATTGGTGC GGGCTCTTGA GTGACAGAAG GCAAAATTAT	2100
CCCGCCAAAT GTTTTAGCAT TTGGTCGACC CGCGAAGGTT ATCCGTCCGT TAACGGCAGC	2160
AGAAATAGCC AGTAATCATA CGAACGCCTT ACATTATATT CAACGAGCAG AAGAATTTG	2220
CCAACAAAAA TACCCGAAAC TAACATAGAA TTAGGAGGAA AATGATGAAG AAAATTAAAG	2280
TAATGACGGT TTTTGGAACA CGGCCAGAAG CAATCAAAAT GGCTCCCTTA ATCAAAGTTT	2340
TAGAAGAGCA AAGTCAAGGA TTTGACTCTG TAGTAGTCGT AACAGCGCAA CATCGCCAAA	2400
TGTTGGATCA AGTGTTAGAA GATTTTCAAA TTACACCGAA TCATGATTTA AATATCATGA	2460
AAGATGGGCA AACCTTAACA GATATTACAA GTCGTGTTTT GAATTTATTA ACTGAGGTTT	2520
TTGTGACAGA ACAGCCGGAT ATTGTCTTGG TTCATGGCGA TACAACCAGC TCATTTGCGG	2580



CTGCATTGGC	TGCTTTTTAT	CAACAGATTC	CAGTGGGACA	TGTGGAAGCA	GGATTGCGGA	2640
CCTGGCAAAA	ATATTCTCCG	TTTCCGGAAG	AAATGAATCG	TCAACTGGTA	GACGTGTAA	2700
CTGATATTTA	TTTTGCCCCA	ACCACTCAAA	GTAAGGACAA	TTAATCAAA	GAAAATCATC	2760
CAGAAGAACA	TATTTATGTG	ACTGGCAATA	CAGCAATTGA	CGCGATGGCG	TAGACCGTTG	2820
ATGCCCCACTA	TCAAAATGAT	TTATTGGAAA	AAATCCCTAC	AGACCAACGC	ATCGTGCTAA	2880
TTACAATGCA	TCGTCTGTAA	AATTTAGGCC	TGCCAATGGC	AAATGTGTTT	AAAGCTGTTC	2940
GTCGACTTGT	AATGGAACAT	CCAGAAATTG	AAGTCATTTT	CCCTATGCAC	AAAAATCCCA	3000
AAGTCCGTGA	AATAGTAGCG	GAACACTTAG	GTGAATTAGC	TCGTGTGCAT	TTAATTGAAC	3060
CATTGGATGT	CAAAGATTTT	CAAAATTTTG	CAGCGAAAAG	TGCTTTAATT	TTAACTGATT	3120
CTGGCGGTGT	GCAAGAAGAA	GCTCCTTCCT	TAGGTGTGCC	AGFTTTAGTT	TTACGAGATA	3180
CAACGGAACG	TCCAGAAGGC	GTAACGCGCAG	GAACATTGAA	ATTGGTTGGT	ACAGACGAAC	3240
AAGTAGTCTA	TCAAGAGGCG	AAAACATTAT	TGACTGATTC	TGATGCTTAC	CACGCGATGG	3300
CGCATGCACA	AAATCCGTAT	GGGGATGGTC	AAGCTAGCCG	CCGAATCGTT	GAAGCCATTG	3360
CCTATGAAAT	GCAGAAATCA	GATAAAAAAC	CAGATACGTT	TACAGCAAAA	TAAATAAAAA	3420
GGCAGAGAAG	CGCTGAAAAA	TGTGCTTCTT	TTGCGTATCT	TATTTTGAAG	AATCAATTAG	3480
ACAAAAAATG	CTAGAAAGTA	GTGAAGTGAA	GATGTATCAA	ACGATTTTAT	TTGATTTAGA	3540
TGGCACAATT	ACGGATTCTG	GTAGCGGCAT	TATGCGTTCT	ATTTTATACG	CAACAGAACA	3600
ATTAGGCTGG	CCAGCTCTTA	GCGAGGAGAC	ATTACGTTCT	TTCATAGGGC	CGCCTTTATA	3660
TGAGTCGTTT	TTGCACATGG	CGCCTTCAGC	TGAAGCAGCG	CAACAAGCAG	TGGGTCATTA	3720
TCGAGCGTAT	TATCAGAGAA	AAGGTATGTT	TGAAAATCAT	GTGTATCCAG	GGATCCCAGA	3780
GGTACTAATA	AGGTAAAGG	AAGCGGGAGC	AAAGTTATAC	ATTGCGACTT	CCAAGCCAGA	3840
AGAATTGCT	AAAAAGATTA	TTACTCATTT	TGATTTAGAT	CGTTATTTTA	CAGGCATATA	3900
TGGAGCGAGT	ATGGACGGTC	ATCGTTCTAA	AAAAGCAGAT	GTGATTCAGT	ATGCATTAAC	3960
GGAAGCTCAA	CTAGACCCAA	CCAAAGAAGC	AATTATCATG	GTTGGTGATC	GTAATCATGA	4020
CATTCTTGGT	GCTCAACAAA	ATGGCTTAGA	TAGTATAGGT	GTCTTGATG	GATTTGGTGA	4080
AGAAACAGAG	TTACAAGAGG	CGGGAGCGAC	ATTCCTAGTA	CACTCGCCTA	AAGACTTAGG	4140
TGCTATTTTG	CTACGAAATT	CTTAATAGAA	AATGGATTTT	CGATAAATTT	AGGAGGATTT	4200
TAAGATTTAT	TATCGTTTCA	ATAGAAAAGC	GCCATCATAG	GCTTTAAACG	TTATCAATGA	4260
TGAATTTTTT	ATCGGGGAAT	GACGGTCTTT	TTATTAAAAA	CATTTAATTT	TAAAGAGGTT	4320
TCATGTATAA	TAGTTTCAGT	CATAAACGAT	AGAATAGTAA	GGAGGGTTAG	TTTTGGCTAA	4380
TGATAACCAA	AATAACCAAG	ATCCCAAATC	TTCGCTTAGA	GATCAAGTAA	CTGGGTCTTT	4440
AAATGGTAGA	AATGATGGCG	ATCAACCTGA	TTCCTCTGAA	AAAAATGATA	GGTCACCCCA	4500
ACCTTCATCT	GATGAATCAC	AAGAAACAGC	GAGCCGCACA	ACACAAACAC	GTGCAGGTTC	4560

TAGAGCTGCT CGAAGACGTG GGAAAGACAA GACAACACAG ACAGTCGAAG AGCCAACCCC	4620
GATTGAAACA GACGAAAAGC CAACAAATAC AAAAAACAA ACACGTAAAA AAGAAGATCG	4680
TTTAGTCGGT CGTATCGTGT TAATTGTAGT TTCTGTATTA GTTTTAATGA TGGCAATCTT	4740
TGGGTTTACT TTTTATAAAT ATGTAGATGC TGGGCTACAG CCACTAGACA AGAACAATAA	4800
GAAGCTTGTT CAAGTGCATA TCCCAGAAGG ATCTTCTAAT AAGCAAATTG CTGCGGTTTT	4860
AGAAGAAAGC AACGTCATTA AAAGCGGAAT GGTTTTTAAC TACTACGTAA AATTTAAAAA	4920
CTTAACAGAT TTTCAAGCGG GCTATTATCA AATGTCTCCA AGCATGACAC TCGATGAAAT	4980
TGGTGAAATG CTAAAAGAAG GAGGTACGCC AGAACCAACT AAAATTGCAG ATGGCAAAGT	5040
AACGATTCCT GAAGGCTACG ATATAGATAA AATTGGCGAA GCTATTGAAA AAAATACAGA	5100
TTTCAAGAAA GCAGATTTC AATGCATTAAT GAAAAACGAA GACTTCTTCA ATCAAATGAA	5160
AGCTAAGTAT CCTGATTTAC TTGAAAGTGC TGCAACTGCA GAAGGTGTTT GTTATCGCTT	5220
AGAAGGCTAT TTGTTCCCGG CAACTTATGA TTACTACAAA AAAGCAACAC TTCCTGAATT	5280
TGTCGAACAA ATGATTGCTA AAATGAACAC AGTAATGGAA CAATATACGC CGACAATTCA	5340
TGCAAAGAAT TTAACCAATC AGCAAGTCTT AACGTTAGCT TCATTGGTTG AAAAAGAAGG	5400
AGTCAAAGAA GCAGACCGCA AACAAATTGC GCAAGTATTC TTTAACCGCT TGGCTGCCGA	5460
TATGCCAATC CAATCAGATA TTTCTATTTT GTATGCTCTA GGAGAACACA AAGAACTGT	5520
AACGTATGCT GACTTGGAAG TGGACTCGTC TTATAACTTG TATAAAAATA CTGGTTATGG	5580
ACCAGGACCA TTAGATAGCC CAAGTGAGGA GTCAATTAAA GCAGTCTTGA ATCCGACACC	5640
AAGTGAATAT TTATACTTTg TAGCAGATAT TTCTACAGGA AAAGTCTACT TCTCAAAAAC	5700
ATATGAAGAA CATCAAGTAT TAGTAGATCA GTATGTCAAT AATTCAAGTA GTGAATAAAA	5760
TGAAACATA AAAGAGCTTG GTGAATCAAT TTTGTATGAT TCACCAAGCT TTTAGATACT	5820
CAAAAAGTAA TCTGTTTAAC AGAGGGAAAG AGTAGCAACT AGAAATTTAT ATAAATAAAA	5880
TTTAGCAAAA TCACAAAAAT TCAATAATTA TATTTACAAT TTAGAAAAAG CGTGGTACTC	5940
TTTTGTGGAT TGAAAGCGGA ATTTTTTTTCA ATAACCAACC AAGAACTACG TATCGTTAAG	6000
CGAAAAAGGA GAATTAACAT GGTAGAAAAA GTATATCCAA TGACTCTTGA AGGAAAAGAA	6060
AAATTAGAAC AAGAGCTAGA AGAATTAAAA ACAATTAAAC GTAAAGAAAT CGTAGAGCGT	6120
ATTAATTCG CTCGTAGCTT TGGGGATCTT TCAGAGAACT CAGAGTATGA ATCAGCTAAA	6180
GATGAACAAG CTTTCGTTGA AGGACGTATT ACAACGTTAG AAAATATGAT TCGTTTTGCA	6240
GAAATCATCG ATAATGATGG CGTGGACCAA GACGAAnTTT CCATCGGTAA AACAGTTACT	6300
TTCCAAGAAT TGCCTGATGG AGAAGAAGAG GAATACACAA TCGTAGGAAG TGCCGAAGCC	6360
GATCCATTTT CTGGTAAAAAT TTCTAACGAT TCACCGATTG CGCAACCnTT AATTGGTAAA	6420
CAATTAGGCG ACCAAGTGAT TATTGCAACA CCTGGTGGCG ACATGACAGT TAAAATTACA	6480
AAAGTTGAAG CAAACTAAAT AATTGAGTAG GGCAGGCAGC TTCCAAGGCC tGTCCTTTTCG	6540

CAATCTATGA CATTTAGGGC TAAAGTCCTA TGGTAACCTG aGGCGCAAAC CACTATGATT	6600
AAATAAAATA AACGAGTAAA ATAAATGACA GAGGGAGGAG GCAAATCTAT GAATAACCCT	6660
TGGCGCTTTT TTATCGTCGC AGAAGCATT ATTTTTATTC TGGCGTTATG GCAAATTGTA	6720
CATAATCCTG GaTTAGCTGT TtAtTAACA ATTGGCGTTT TACTTGTGGC CTACGTTTCC	6780
AGGAAAGCAT CTAAACACA TTTTAACAAC TTTCAATTCG TCCTCGGCGT TGTTTTTATT	6840
GTCATTGGTG CAATGAATAG CACGGCTGTT TGGTTTATGT TGATTTTGG CGTACTCTTT	6900
ATCGGCTTAA AAGGCTTTGA GATTTCAAGC GTGGATATAG CTGAGCGAGC ACCTTGCGA	6960
AAAAACAAA TGATTATGGT GGAGACGGCG GCAAAGAAC CTAAAAATGG CAAACGGTTT	7020
AAACGCCGCT GGTtTGCCAA CGAACGCATT GGTAACAATA TCTATGAATG GGACGATATC	7080
AATATTGATT TAATCTCTGG GGACACCATT ATTGATTAG GTAATACGCT ACTACCGAAA	7140
GAAGACAATA TTATTATTAT TCGTAAAGGT TTTGGCCGCA CGCGAATTCT AGTGCCGTTA	7200
GgTGTAGCTA TTTTGTTAGA ACATTCAACT TTTTACGGAA CGGTACGTTT TGAAGAAGAA	7260
AAATATCAAT TGAAAAACGA ATCATTAATA ATTTACAGCA ATGATTATGA TACCAATCTT	7320
CGTCGTTTGA AAATTATGAC GAACACTTTA GTAGGAGATG TTGAGGTGAT CCGTGTATGA	7380
CCGATCGGAT TTCAAGACGC ATGATTTCA TATATGCGTC CCTTAGCACC TTTATTGTTA	7440
TCTTAATTAC ATTGTTTCA TATTTTCATT CGATTAAACA AAACCGGTGG TTATTAGAGC	7500
TTCTTCAGAG AAAAGCTTT TATTTACCAC TAATTGTGCA CATTGTTCTC ATATCCTTAC	7560
TAATAGGCTT ATTGACCTTT TTA CTGATTT CATTGGTTCA AAAAGGGCAA TATGGACGGA	7620
TTGAAGAAAA ACTTCGGTTA TTGGCCAACG GTAATTATGA AAGTCCAGTC TTAAACAAAC	7680
CAACGACCAG TGAAATCAA GACCATTATC TAACCGAAGT CGAACAAGAT ATTTGGTCGA	7740
TTAAAAATAA ATTATTAGAG ATGTCTAAAG AATTGCAATT ATTAACAGT CGACCGCAAT	7800
TAATGGATGG GCAAACAAA GAAGAAATTT TAGAGAACGA GCGGCATCGT TTGGCGGGG	7860
AGTTGCATGA TTCAGTCAGT CAACAACTTT TTGCAGCCAT GATGATGTTG TCTGCATTAA	7920
ATGAACAAGC ACAACGAACA GAAACCCCGG AACCATATCG TAAACAATA GCCATGGTGG	7980
CAGAAATCAT TAATGCCTCC CAATCGGAAA TGCGCGCGCT ACTATTGCAC TTGCGTCCTA	8040
TCAGTCTAGA AGGAAAAAGT TTGCGTAAAG GTATTGAACA ATTACTGAAA GAACTACAAA	8100
CAAAAATTAA AATTGAATTG ATTTGGGATG TTGAAGATGT TCATTTAAAT AGCAGCATTG	8160
AGGATCATCT TTTCCGAATT GTGCAGGAGT TACTTTCAA TACCTTAAGA CATGCCAAAG	8220
CAAAGGAATT AGAGGTATAC TTACACCAAG TCGATAAAAA CGTGTTATTG CGTATTGTTG	8280
ATGATGGTGT CGGCTTTGAT ATGAAGGAAC AAAGTAATA AGCCGGTAGT TATGGCTTAA	8340
ATAATATTCG AGAACGTGTT GTCGSCATGG GCGGTACAGT TAAAATTATT AGTTTTAAAG	8400
GGCAGGGAAC CAGCGTTGAA ATTAAAGTTC CTGTCATAAA GGAGGAACT GCAAGTGATC	8460
AAAGTAATGT TAGTGGATGA CCATGAAATG GTCCGTTTAG GCGTTTCATC ATATTTATCT	8520

ATTCAGAGG ATATAGAAGT CGTAGGCGAA gcAGAAAACG GTAAGATTGG CTATGAAAA	8580
GCATTGGAAC TACGTCCAGA TGTTATTTTG ATGGATTGG TAATGGAAGA AATGGACGGC	8640
ATTGATTCAA CAAAAGCGAT CTTGAAAGAT TGGCCAGAAG CCAAGATTAT TATTGTGACG	8700
AGTTTTATTG ATGATGAAAA AGTGATCCG GCGATTGAAG CTGGTGCAGC GGGCTACCTA	8760
TTAAAGACAT CAACAGCACA TGAGATTGCT GATGCAATTC GGGCGACTTA TCGCGGAGAG	8820
CGTGTGTTGG AACCTGAAGT GACGCATAAG ATGATGGAAC GGTTAACAAA AAAACAAGAG	8880
CCGGTGTTGC ACGAAGATT GACAAACCGG GAACACGAAA TTTTAATGTT GATTGCACAA	8940
GGTAAAAGTA ATCAGGAAAT AGCTGATGAA CTCTTTATCA CTTTGAAAAC AGTTAAAACA	9000
CATGTTTCAA ACATTTTAGC AAACTAGAT GTGGATGATC GGACCCAAGC GGCGATTAT	9060
GCTTTTCAAC ATGGTTTAGC CAAATAAATA TCAGATAAAT CCATAAATAA ATTCTAGGAA	9120
ATCCTTCTAG GAATTTATTT ATGGATTTGT TATACTAATT TTAACGAACA GACAGGAGCG	9180
GTCAGAAGAA ATGAAACAAA ATTTTGCAAT TATTGGTTTA GGACGTTTGG GCGGCAGCAT	9240
ATGCCAAACT TTGGTGGAAG CTGGTCAAGA AGTTTGGCC ATTGACAGCA GTGAAGACCG	9300
TGTCATGAA TATATGAATA TTGCAACGCA TGCAGTGGTG gcTAACGCCC AAGATGAAAT	9360
GACGTTGCGT TCTTTAGGGA TTCGGAATTT TGATCACGTA GTGkTGGCAA TTGGTGAAGA	9420
TATTCAAGCG AGTATTTTAG TGACACTAAT GGTAAAGAG ATGGGGGTTT CCAATGTCTT	9480
GGCCAAAGCA GTCAATGAAT ATCATGCCCC TGTCTTAGAT AAAATTGGGG CAGACATGGT	9540
CGTTCACCCA GAGCGGGATA TGGGCATTCG AATTGCTCAC AACTTGTTT CACGCAATAT	9600
TTTAGATTAT ATTGAATTAT CTAGCGAATT TTCATTAGCT GAGGTTCTGG TCACCAATCC	9660
TAAATTTTAC AATAAACCT TAGCAGAATT AAATTCCGT CAGCGATTG GTTTAACTGT	9720
GGTAGCGATT CGCCGCTCGA AACAGAAAGT AATTGCTTCG CCAGATGCTT CTGAAATTGT	9780
TCGAGAAAAC GACAATTTAT TAGTTATCGG TGATACCGCA GACGTGGATT TATTAGATGA	9840
AAAAATGAAT GGCTAAGCAC TTTATTCTCT AAAAGCACTG CGCTGTGCTA CAATAAATGG	9900
CGACAGAGAT GTGAAGGAGG AACAAAGAAT GAAATTAGAA GTGACACCGA AAGCCCAACA	9960
ATGGTTTAAA GAAGAAACCG GGGTTCAGCC TGAGACAGGC ATTCGTTTTT TCGGAAAAAT	10020
TTACGGAAAA ACACCTGTTC ATGATGGTTT TTCCATTGCG ATGTCTGTAG AAGCGCCAGA	10080
TGAGCCAATG ATTAAAGAGA ACCTAGATGG GATTACCTAT TTTATTGaAG AAACAGATGA	10140
TTGGTTCTTT AAAGGGTATG ATTTGATTGT TGATTATGAT GAAAAGAATG ATGAACCGCG	10200
TTATCmATTT GAAGCAAACC mAGrAGaTTT AACmAAATAA TTTTAGTGCT AAACAAAAGA	10260
GGCTGAAGTA GTGCGTAGGA CTATTCTGTG TGCTTTCGTT TAACCGCTAG TCAAnnGTGAG	10320
AAAAAATCGT TTTGGATTTT TTCTCAGGTC TTTCTATTTA AGTATTTTAG TTGGATATCG	10380
CTATTTAATA A	10391

(2) INFORMATION FOR SEQ ID NO: 134:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3852 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

AATGTTAAAA GGAnAAATAA TAGAGTTAAA GTTGAAAAAG tGgAGTTAGA TTATGTAGGA	60
AGTATTAmTG TkGATAkGGA TTTATTAGAA GcAGsAGGaa TATATGAATA TGAAAnAGTA	120
CAAATAGTTG ATGTTAATAA TGGAAATAGA TTTGAAACTT ATACAATAGC TGGAGAGAGA	180
GGAACAGGGA TGATTTGCTT GAATGGAGCT GCAGCTAGAT GTGTATCAAC TGGAGATAAA	240
ATAATTTTAA TGGCATATTG TGAATTAAAT GAAAATGAGG TTAAGGATCA TAGACCTAAG	300
GTGTTTTTTC TCGATGATAA TAATAAGTT GAAAGAGTAA CATCATATGA AAAACATGGA	360
AGGTTATCAG aTATAATTCT TTAGGTTTTG TAATGTTATT TATATATAGC TAGACTATTA	420
TTGAAATATT TATTGATTAC TAATTGGATA GATGTCAATA AAGTAGACAC AAAAAGAGAA	480
GATACTTAAA GATTTTCAAA GTAAAGAGCC TCCATTTTCAT TTGGTGtAAG CATATCTCAT	540
GTTCCATGAG GACGTTTGGT GTTATAGAAG CCTTCAATGT AGGCAAAACA AGCCAACCGT	600
ACTTCTTCAA TTGAATGGAA TGAACGGCGA TTTAATTCTT CTTTTTTCAT ATATTAAAAA	660
AAAGCTTCGG TAACGGCGTT ATCCCACGGA TAACCTGGTT TTGAAAGTGA GTGGACAATC	720
GGATGATCCT CTATAAATTT TCTAAATTCA AAAGAGGTAT ATTGAGACCC TTGATCTGTA	780
TGAAAAAGAA CAGAACTCGT TAAtGcgAAT GGCAACAGTT TTTTGGACAA ATTTTATAAG	840
GTGCAGAACT TCTTTCCGTA TACTATTCCG ATTGTCCTAG ACAATGAGCC TCCTCGGATG	900
TGACGATCTT CGGCAGGAGC TAACAGTTAA AGTTAGACTG CCTCGCTAGC GTTAgcACAT	960
CCGAGCTCAT TATCAAGGAT TCGcTGCGCC AATCTCTGGT TACGGTAACC AACCGGTGTC	1020
TCGTAGCCAA GTGTACCGTG CAACCGAAGt GGnTTCCACC AATTGACATA GTCAAATAAC	1080
TCCAAATCCA ATTGTTGTAA GGTTCCAAT GTGTATTGAT AGACAAATTC TACTTTCAAC	1140
GACTTATAAG TTGATTACAG TACGGCATT TCAAAAGGAC AGCCTTTATG ACTCAATGAT	1200
CGATTGATGT CAAAAGTCGT TAATAATTCA TCAATAGCTT GGTATCAAA CTCTTTTCCA	1260
CGATCAGTAT GAAAAATCTC AACCTCTGTC AGAGGTTGTC TGATACGGCT AAATGCTTTT	1320
TTTACTAGAA CGGCATCTTT ATGTTCTCCA CAAGAATAGC CGAGAATTC TCGATTGAAC	1380
AGATCCAAAA TGAAACAGAC ATAATTCCAT TTTTCCCGA CTCGTACATA AGTCAAGTCT	1440
GTTACGATCG CTTCTAATGG GTTGTCTCTT AAGAATTTAC GATTCAATAC GTTTGTCGTT	1500
TTGGCTTCAT TGCAAGTAGA ATGATGTACT TTAAAATAAG CAACAGTATA GCTCGATTTT	1560
AATCCTCTAT TTTTCATGAT TCTACTAATT TTTCGTCGGC TGATCTGAAT GCCTCGTTTT	1620
GATAAGGCTT TTTTATTTT TCTTGAGCCG TAGgCtTTTCG GCTGAGGATA AATTCTTCAG	1680

CGACTACTTC	TTCAAGTTCT	GATTTCGTCTT	TCTTTGGTTT	TGATTGATAA	TAATAGGACT	1740
GACGTGATAG	ACCTAATATT	CTGCACATCG	CTGATATAGG	GTAAAGATGC	TTATTCGCAT	1800
CGATTACTTG	TCTCTTCGTC	CGAATATCAG	CGCTGCTTGC	TTTAAAATAT	CATTTTCCAT	1860
TTCTAATTGC	TGGTTTCTTT	TACGTAGTTC	TAACAATTCT	TTTTGTTCAG	GCGTAAGATT	1920
ATCTTTTTCT	TTGAATGAAC	CACTCGTTTT	AGATTGCTTT	ACCCATTGTG	CAAATGCTGA	1980
AGCCGTTAGT	TCATATTCTC	GAATGATTTT	TACACGTGGC	TTCCAGCTA	AGTAAATATT	2040
GACTATTTGT	TGCTTGAATT	CTTGTGAAAA	AGTTCTTCGT	GTTCTCTTAG	ATATAAAAAAT	2100
TCCTCCTGGT	ATGtTTTCTT	CTAGTCTACA	CACCTTAATT	TTTCTGTCTA	GTTAATTGTA	2160
GCCTATCCAG	TGCTTCTGCC	GGAGAAAATC	AGATCCGAGG	ATGCTCATTG	TCAAGGGCAA	2220
TCGGAGCCAT	AACACCGCAG	CATTCACAAC	ACCTTATAAT	TTTTGTCAAA	AAAAGTGTG	2280
CCATTCCnAT	AGATGAAATT	CTTCAATTGA	GAGTGTA AAA	TATTTGTGT	AAATAGAAAA	2340
AAGGAAGTCC	CTTCTGTAGA	ATAGAGTTAC	CACAACACAT	TCACAGAAAA	GAGGACTTCC	2400
ATATGAACGA	TTTTACTACA	GAAATTCTAA	AGACyCTAGC	GAACnAAAGG	CGATTGgATG	2460
AATTATTAGT	TGCTTTAATT	ATTTATAAAT	AACTATAATG	CTTTTATTAT	AGTATCGTAA	2520
ATAGTTTATA	TGAATACATT	AATTGTAAAA	ATGGTGGATG	GCTCTGAACG	ATTAACAATA	2580
TTTTTTTGT	TTGACTATCA	TTTATTGTT	CAGAAATAAT	TTATTGCACT	TTTTGAGGTA	2640
AATAGTTTAG	AATTATTTTA	ATATAAGTAC	TTCAGATTTA	ATAAAGGAGA	AATATCTAAT	2700
GAAAAAAATT	GATTTAATAA	ATATGATAGG	TATGTTAATA	GGGATATTAG	TAAATATAGT	2760
AATTTTCACA	GATTGGCTCG	GGGTGTTATT	TTCTAATCTT	ATTCCCATAC	TTATTATAGG	2820
CATTTGTGGA	ATTATACTTT	CAATATTAGA	ACTTTTTGAA	AGCAGAAATA	CAATGAATAG	2880
AATATTTGCT	TGTATTATAT	TAATTGTTAA	TCTATTGCCT	ATGGTCTATT	TTACTTTTTT	2940
ATATTTTGCA	CTGGGATGAA	ATAAATTAAC	CAAAAGTTTC	TTTTTCTTAT	AATGTGTATC	3000
AATATTAAAA	CTGAAAAAGA	TATGGATAAT	AGGATCAAAG	CCCTCATTGG	AACACTCAAT	3060
ATATTATCAG	ATTTAAATTT	CAAATAGATA	ATTTCATTTT	TTAGTAATAT	CGAAAGATAC	3120
TTTTTTAATG	ATGAGAGTGT	AATGTAAACC	GTGTAAGTAA	CCATTAATCC	TAGGCAGGTG	3180
GTTGcTTCGC	GGTTTTTCTT	TTTCTACAAT	AAGATCAAGA	GGTGATTACA	CGATGGCTAG	3240
AAAGAAAAGA	AATCCTGATG	CCGAAAAGTT	AGCTGAATCC	ATTCTGAATG	CCTATCAACC	3300
TGAATCTGTC	GATGACATGC	AAGATGCTTT	GAAAGATGTG	TTTGGGCCCC	TTTTTGAAAA	3360
AATGCTTCAA	GGAGAATTGA	ATAATCATTT	AGGTTATGAT	GCCCATTCTA	AAGAGCCTAA	3420
GGAACACGAT	AACCGTCGAA	ATGGCTATGG	AACTAAAACG	CTTAAAACCA	GTTTTGGTGA	3480
AGTAGCTATT	GATGTTCCCTA	GAGACCGGGA	AGCTTCCTTT	GAACCAGAGT	TAATTCCTAA	3540
GAGAAAACGA	GATGTCTCCG	ACATCGAAGG	GAAGGTCTTT	TCCATGTATG	CACGAGGAAT	3600
GAGTCAACGA	GATATTGCCG	CAACCGTCGA	AGCTATCTAT	GGCTTTGATA	TTTCCCATGA	3660

AATGATTTCA GATATCACTG ACGCTGTCCT TCCTGAATTG GAAAGAATGGC AAGCCCGCCC	3720
CCTAGCAAAG TGTTATGCCT TTCTATTCGT TGATTGTATG TATGTTACTT TACGAGAAAA	3780
TTATGAAGCC AAAGAATATG CTGTATACAC CATTCTTGGC TATGATCTCA AAGGAATAAG	3840
GAGATTTTAG GA	3852

## (2) INFORMATION FOR SEQ ID NO: 135:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8805 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

ATTGGCTAAG CGTTTAAACA AATGCGTCTT CTAATAAya GGTTTTAATG ACATGACTTA	60
nTTCyCCTTT TtGtCCATAA TTTCTTGTA tCGTgCACGA TTTTGTTTCAT TTCTTCTTGA	120
TTGACAGAAC GAATGTCAAA TTCAATGATG CcATTGTTGG CTTGATATTC TCGCGTATAA	180
ATGGCTGGAC TTTCTGCTTT TAATGCTTGG ATGACTTCCT TAGCGGTTTT TCTGCCGTCC	240
ACTTTGACAC TGGCGCGATA AATGTCTCGG CCAGCGCCAT CTTGGACGAT TTTGGCGGTT	300
AAATCTGATA AGTTGTTGAT GGCTTCAACA AACGGTTTTA AGCGCTCTTG CATCGAAGCC	360
CCTGATTCAC TGCCATGTGC CAAGTATTCT TCCACTGCTT GCGTAAAGCC CAAGATATTG	420
TCTTTGCCAA TTTTCATTGC TCGACCAATG CCTTTGCCTT GTAGGCGCAC CCAGTCAATG	480
TATTCTTTTT TCCCGACCAC TAAACCGGCA CTTGGCCCTT CAATTGCTTT GGCGCCGCTG	540
TAAATCACCA AATCAGCGCC TGCTTCAGTG TATTTGAAAA GATCTTCTTC TGcTGCCGCG	600
TCAACAATTA GCGGCACCTT GTGGCGTTGC GCCACTTTGG CCGCTTCGGC CACCGTCAGC	660
ATACTTTTTT GCACGGTATG GTGGCTTTTG ATGTATAGAA TGGCCGCTGT TTTTTCAGAA	720
ATCATCATTT CCACGTGTTC TGGACTGCAC ATGTTGgCGT AGCCAGCTTC CACCACTTGG	780
CCGCCACCTT GCGCTACCAT GACTTCTACA GCGTGCCGT AGTCCACGTT GTGCCCTTTA	840
GGTAAACAA TCTCCCGCTG TTCGATTTTT TCTGTATAGG GATGATACGC GTGGTACAAG	900
CTGCCTTTTC CAATCAAGGC AGCCACCGAT TGGGCAATCC CCGCCGAAGC CGAGGAAACG	960
ATTTGGGCAT CTTCTACCTT CAACAAGTTG GCTAAAAAGG CCCCTGTTTG CACACTAAGT	1020
TCACTCATTT CAAAGAAATG TTCGCCGCCG AATCGTTGAG CTGCTAGGAC TGCCTCCGAA	1080
ACTTTGGAGA CACCTAGAAT GGTCATTTTG CCAGAAGCAT TAATGACTTC TTTTAAGTGG	1140
AATTTTtCGT AACTAATTGT CATAGATTG TCCTCCAATA ATTGTTTTGA TTGGGCGAAT	1200
TTGCTCTTTG GCTACTCGTG TCAATCCATT AGAATCAGTC AATGTTTTTT CTCTGCTTG	1260
AATCGTGAAA ATCGTTAAAT CAGCATCTTT ACCTATTTCC AACGTGCCTT TTTGCGTTAA	1320
ATGGAAATTC TCAGCAGGAG CCTTGATTAC CTTTCAATA ATTTCTGGCC AATCATAGCC	1380
GACCACACGG AGTTTTTCCA TCGTTGTGGC TAAATCATAA ACAGGACCAT TTTCTCGGTT	1440

ACGGATGTAG	ATATCTGTGC	TAATTGATGC	AGCTTTCATT	CCCTCGCGCA	ACGCTGTTTC	1500
AGCTACGTGA	AAATTGAAAC	TATCCGTACC	ATGCCCGATA	TCAAATACGA	CTCCTTTGTT	1560
GTACGCTTGC	CAAGCGAAAT	CTTTTATTTT	ATCCGTTGCT	TGGTCTAAGA	TACCATTTTC	1620
TTTCCCATTA	AAACAATGGG	TTAGGACATC	CCCTTTTTTC	ATCAAGGCTA	AAATTTTCATC	1680
TAAATGAGGC	GGTGCTGATC	CAATGTGAAC	CATTAAGGGA	ATTTCTTGGT	TCTCTTGCTG	1740
AATCTGTTTG	GCCAATTCTA	AAGGCGTAAT	GCCGTTATCG	CCAATAACGG	TCCTACTCAT	1800
ACGGGCTTTA	ATCCCTACAA	CAAAATCTGG	TAATTCTTGA	ATCGCTTTTT	TGACTAAACT	1860
CGCTTGTA	TTACTTAAAT	CTGCGAGTTC	GTCCTGAGCA	ACGATGCCCC	ATTTAGAAAT	1920
ATTGACTAAT	CCAAAAACAT	TTGTTTTTGC	TTGcTGCGCT	AAGTCATAAA	ATTCATGAAT	1980
GTTCCTCAGCA	CCTGTTGTCC	CAGCATCAAT	CACTGTGCTA	ACACCCTTTT	TGACCCCAAT	2040
TTCATCTGGA	TAATCATAAT	AAAGAGCCAT	TTTTTCAAAA	CAATGAACGT	GATCATCTAT	2100
CCAGCCTGCG	GATACATAAG	TACCTGGTTC	TAAGTGGATA	GTTTCTTTTG	CAGAACCTGA	2160
AATAGTTGCT	GCAACAGCAG	CTATTTTTTT	CTCTTTAATT	GCAATTTCAA	CAGGCATACC	2220
ATTTACTGTC	TGTCCATTTT	TTATAAGTAA	ATCGTAGTCC	ATATTCTTCC	CTACTTTTCT	2280
ATAAATCTAA	ATTTTGTCCA	AGGCTTAATG	AATTCTAATA	ATACAAGTTC	TTCTTCTATA	2340
ATACGCCCTA	CTCGGTTTTT	GCGTGAGTCT	TGATGTGGTT	CTAAAACGAC	TTGTAATTCA	2400
TTTTTGATTT	TTCCAAATGC	ATCATTTCCT	ACAACAACAT	CTCCAGGTTG	TAGCATTGCT	2460
TGATTATCAT	GTGGGGGATT	AACTTCATTT	TTATATTTTT	TACGCACTTC	TGTGGAACGA	2520
ATCATCTGCT	GAGTGATATC	TCCACGGCGA	AAATGTTGAT	TTTTTGTTAC	GATTTCTTTT	2580
TCAATTTTCAG	AAGTCGCTTC	TTCAAAAACA	ATGGTCAATT	CTGTTTGATA	ACGATTCAAC	2640
TGACCCAATG	CTTCTAATTC	TTGTTGGAA	GCATAAGCAT	TTCCAATAAC	AACATCATCA	2700
ATTAAATTTG	TTGCAATAAA	ATGTTTAGCT	TGAACAGTGA	CTGGCAACTG	TCTATGCATT	2760
TCTAAAGTAG	GTAGCCCATC	ATTCACATCC	CAAGGACCAA	TTTCACCTAC	CTGTGATGAA	2820
ATAAAAGCAG	CTGTCCGAAT	ACCTTCCTTT	TTAAAACGGA	TGCTACATTT	TTCAAAGAAA	2880
TCATACGGCA	ATGCAGTTCC	CGCTTGTTGA	TAAAAGTTAT	GACAACCATA	TAAGAAAGGT	2940
CGATTCGCTT	CATACGTCAA	AATGTTATCA	AGATAAGCGA	CATCATTACT	CATATTCAAC	3000
TCAATAGCCA	CGCCGAATGG	ATTAAACGTC	AACATGGCTT	CTTTATTTCC	ATCAAACCCT	3060
GTATCCAACC	GAATCCCATC	AGCTCCTAAT	TCGCTGAAAA	AAGTCAAATC	ATCATATGAA	3120
ATTTCCAATT	CATCAAAGAT	ACTTGAACG	ACATCTAAAA	TTGTTTCAA	TCCTAAATTT	3180
TTGCATAAC	TAATTAATTC	TTTAAATTTT	TTCTGAACCG	CTTCTTTGCC	ATCGGTCACT	3240
TCTAACATAC	TCATAAAAA	TCGAGTAAAC	CCACATTCGC	TGGCTTTTTT	TAAATACGCT	3300
TTATCTTGAT	TGATATCACT	ATGATCTGGA	TAAACTGATA	CACCTAATGC	TCTTTTCATC	3360
TTCGTTCTCA	CACTCCGCTA	TTTCAAATTT	TATTTCGAAA	TTCCAAGCA	ATCACCGAGC	3420



TACAC <sup>Δ</sup> AAGAT	TGTTGTAGAG	CCAACGAGCA	ATACTGGTTG	ATTTAGAAAT	CCGCCAATTA	3480
TCATGGAAAG	AGTAATACAA	ACTGCCGTAA	AAACCAACGT	CTTTACCAA	CCAAATTGAT	3540
AAACAACAGT	AGGCAACTGC	CCTTTTTCTA	TAATAATTTG	ACTAACGATA	AAAAAGAAAA	3600
TGAACCAAAT	AATCATCTTA	TACGCTCCTG	ACTTCTTCT	CTGACTTAGA	AGGGGCAGGA	3660
TATTCATATC	AGAACACCCG	CCCTAACATT	TTTTAATTTT	TTGCTGCTTG	TTCAGCTGCT	3720
AATTCTTCTG	CTTCTTTATC	CAATTGTTGT	TTTTCATAAA	CTTGAAAAA	TGGATAGTAG	3780
ATAGCCATTG	TAATTAAGAA	GTTAAACAATC	ACTAATACTG	CGGCCGAGAA	ACTCCAATTG	3840
GTA <sup>Δ</sup> CTCATCC	AAGCAGCAAT	TGGTGCTGGA	ATTGCAAACG	CTAAGCGTGC	AGCCATCATC	3900
GGAATTATAT	TAGCGACTGT	TAAGAAATAA	GATAACGTTG	TGGTAATTAA	CGGTGCCACG	3960
ATAAACGGAA	TACCTAAAAT	AGGATTCATT	ACGATAGGCG	TTCCAAAAAT	AACTGGTTCA	4020
TTAATATTAA	ACAAGCCTGG	TAA <sup>s</sup> AAC <sup>s</sup> AT	AAACGGCCAA	GACTTTTCAG	ATACGTTGAT	4080
TTAGAAAACA	TCATTAA <sup>Δ</sup> AAC	AACTAAAGAA	AGCGTAGTTC	CAGCGCCACC	AATCCAGATA	4140
AACCATTGTA	AAA <sup>Δ</sup> CTGTTC	TGTAAAAATA	TTAGGTA <sup>Δ</sup> ACT	GATGAACGTT	TGCGCCATTG	4200
GTAAAAGCTT	CAAGGTTTTC	TGCAATCGAC	ATATCCCAGA	AAGGACGAAT	AACTGGGCCC	4260
ATAATCGCTG	GACCATGAAT	CCCTAAAACC	CAGAAGAAAC	AGATTAAGAA	GACTGTTAGT	4320
AAGCCGCCAA	ATAAACTGTT	ACCAGCAAGG	ATACCTTTAA	GTGGCATTAA	TAACGTACTC	4380
AAGAAACCAT	TCAAATCAAA	ACCAATTACG	TGGCGAATAA	CCCAGAATAA	TAAAAGAATA	4440
ACTGCTCCTG	GAATTAACGC	AATAAATGAA	TTTGACACTT	CAGGCGGTAC	GCCATCTGGC	4500
ATCTTAATCA	TAATATCTTT	TTCAATAAAG	AAGCGATAAA	TTTCAACAGA	TAACAGCGCT	4560
GTTACAATCG	CACCAAATAA	AGAAGCTGAA	CCTAAATTTG	CTAAATTAAT	ATAGCGCCCT	4620
GCTGTAATGA	CGTTGTCAAC	ATCTTCAAAG	ACACGTGTTG	GTGGTGCCGC	TGTTACTAAA	4680
AAGGCCATTA	ATGCTAAAAT	TCCGCACGTC	AAAGAGTCAA	GCTTATAACT	TTTGGCCAAT	4740
GATGAAGCAA	TCCCAAATGT	TGCATAAAGT	GCAAGTAACC	CCATCGTATA	GCGGGAAGGA	4800
ATATCTAAAA	TTGCTTGATA	AGGTTCAATC	AATTTTGTGT	ATGCCTCAAT	TGGAATATTT	4860
TGAAAAATTG	TAAAGAATGA	GCCCACAATT	GTTAAAGGCA	TCGTTGCAAT	TAATCCTTTT	4920
CTAATCGCTG	TCATATGGCG	TTGCGAACCA	AATTTATTCG	CAATCGGCAT	TA <sup>Δ</sup> CTTA <sup>Δ</sup> ACT	4980
TCCATCCACT	GTACAAAACC	ATTCA <sup>Δ</sup> TTGT	TAATTCCTCC	GTTCAACTAA	TAA <sup>Δ</sup> ATAAT	5040
AATTATTTTCG	CATTTATAAT	AACATAGTAA	cATTATAATG	TAA <sup>a</sup> CGCTTA	TTCTTATTTG	5100
TGACAGAGAA	AGTGCAAAAC	AAAAGGCTTT	TTCTTATTAT	AAAGAACTG	GG <sup>a</sup> ATATTGC	5160
ACCAAAAAATC	ATTGCGCCTA	AAATTGCGCC	CCCAGCGATT	GGTTTTTTCC	AAATGTAAAA	5220
ACCTAAAGCT	CCTAATGTAG	CACCAATCCC	AATAGGAATC	GAAGCACCCG	CCGCACTCAA	5280
AATAATTAAC	GGGCCTAAGA	AGCGTCCTGA	GGCATTACCA	GCCCCATCA	TCACATCCGC	5340
ACCAAAAGTG	GAATTCCTT	GATTGATAGT	AAATTTCCGA	ATCAACACAA	TTACGCCACC	5400

AACTGCTAAC CCTAGTAGCG CACCTGTTAA AAGTGCTAAG CCAAAAGTTTT CTACAGGTGC	5460
TTTAATACCC ATGCTGAGGA AAATTGCAGG TACGCCAATC CCGATCCCCG TTAAAATTGA	5520
GCCACCTAAA TCTAAAATCC CGACTAAAGA ACCTTCTAAA ATTCGTGCAA ACAGGAAACT	5580
TGCACCAAAT GCTGCGGCAG cGCCATaAGA ACCACCATCT AAACCAGCTT TTAGCATCGC	5640
AACAATTGAA ACTTCATTAA AGGCGCCCAC ACCATAAACT ACATACATAT GTGTTCTCTGC	5700
AAAAATTGCC GCTGACATAA TCCCGACTAA AATGGGAAAG CCCCAATCAG CATACCAGAA	5760
ACTTTTGCCT TTTAAACCTT CTTGAATTTT TTGTTCTTGT TGTGCCATT CATTTTTTTC	5820
TTGCTGTTCC ATCTATGAGA CCTCCTTTTT TCTATTTAAC GGTGAAGACA TTGTGTAAAT	5880
TATTTAACCA TTGAGGAATA GACAAGCGGA AACTTTCGAT CATTTCTAAG TCGAACCTC	5940
GGAAGAAGCC ACTTAAAATA AATAATAAGA TAATGACACC AAGCATGCCG CGTGTCTTC	6000
GGGTCCACCC TAAATCATCG ACACCTTTAC CGATTAAGAT CCCTAATACA ACTCCTGGGA	6060
CAGCATTGCC CATGACTAAT TGAGATAATC CGCCAAATAA TGTCCCCAG AAACCCGTCC	6120
GTTTTCTGC ATCTAAAGCC GCTAGCCAGA ATAAACAGG CATCACAGTA TTGATTAAAA	6180
TCGTTGCTGC AGGTACCAAC ACCGCAACCG CTGTTACTTG TAGCGATTCT GGAATTGCTG	6240
CAGCTGTACT ATTTAAGAAA GCCACAGTCG CCATTCCTAC AAAGACCCCG ACAATCCCCA	6300
TTTTTTTGGG ATTATGCATC GTTTTAGTGA CATCTTTATC CTTTACCAAT AAAATTGCTG	6360
CAGCCCAGTT AGGCACAATC CGGTGGGTCA CATCTTGTGT AAAGGCCCT GCTCCGACGG	6420
CCGAAGCCCA AGCATTAAAG AAAAAACCTA AACCAAAAGA AAAGTGTGAT GCTGGATCTC	6480
CTTGTCGGC GTTCATTCT CCTAATGTTT TAAATGCCCC TAACCCTTGC GTTTGTGGTG	6540
CATGAAACAT CCGAGCAGCG CCTGCGCCAG CTGCAAAACC TAGTAAACCA CCAATAAGCA	6600
ATGACTTCAG CAAAATGATT ACAAATTCCA TTCCTTCTTC CTCCCTTTAT ACGCCTCTTT	6660
TTGAAAGAAA CGGAATACT ACTGCGTTCG GTTCTTTAAT TGTTCCGTT AAAAAATCAA	6720
CTTCTTCGAC TGCAATGGCT TGGATTCCA CTTCTACAGC TAGTTCGACA TGATAAATCA	6780
CTCTCTTTCT AGGAAAGAAG AAGAACAGAA ATTTTCTGT ATAACTTGT TCTTTGCCT	6840
GACGAATAGT TACTTCTTC GGTTCATCC GGATGAGGAC CTCTGTTGTT TCTTTAATA	6900
CTTTTTTTGA AACATCATTC AATGCACTG CAAAAGCTTC ATTCTTCTTA TTACCTTTTC	6960
CAGAAACAAT CACGGTTTGT GTTGACCGT TTGTGGCATC CATAATTACC GTCCTTTCT	7020
AAGTTACTAC TTTCTATGCT CCGTGTCTT TAATAAATGC TTCAGTTACT CGTTTACCAA	7080
GTTCTTCCTG ATCCATAAAA CCAAATCCAA GAACAGTTT CCCATCGTTA ATTGCTGTTA	7140
CCCCTTCTTC GATTGATCGC ATGCCATGTC TTTCGGGATA CCCGTATTTA GTAGCAGCCG	7200
TTAACGCTCC TGCACCACCA CTTCCACAGA AAGAAATGCC TAAATCAGCC TTTCTTGTT	7260
GCATCACATC TCCTAAACGC ATATCAGCGC CCATTCCTGG AACAACAACC GCGCGACCAC	7320
CGGCTTGCTC CACTCCTTTT GCAACATTTT GTCCTTTTCC TAAACGATCT GCGATAACTA	7380

CTGTAATCAT AACTTTTCCT CCTATTAATT GTTTTGTTTA GCAACTTCAA AATGAACAGC	7440
TAAACATAC CATTATCTT CTGATACTG ACCAATTGCT TCTGTTGTTT CTTCTGCAAT	7500
TTTCAACGTT TCTGGCGTTA GTTCCGTAAA TAACGTCCGA TCAACACCAG CTAATGTCGT	7560
TCCTTCTCTT GAACGATTCA GCATCTCATT GAGATGGTTA ATCAAAATCG TCCACTGTAA	7620
TTCCGTTGGT TCAATTTGTG CTTCTGCTAA TCGCGCACTG GTTGTTTGA TAATCGATTG	7680
TAGCTTTTCT GGTTTCAGCAC TTTGTGCAAT GATCTCCTGT GCTTCCATTG CTAGTTCCAT	7740
CCTCTTCCTC CTTTGTTGT TCTAAATAAT TAAGTAATGC TGAAATTTCA GCATGATTGA	7800
CAGGTAATC ATTTTTAGCA AATAAACATT CAATTTGCCG AATTTCTTCT TGCCTTTCAT	7860
GAGATACTTC TCTATTACAA GGAATCTCTT CGTATTGTTT ATTAAACGTA ATCCGATGTG	7920
TCATTAATAA AACGTGAAGT AAAAAGGCTT CTTGATATTC GCTCAGTAAC TTTGTTTGAA	7980
ACATTCGTTT TAACTCTTCA TAAATAATAT ACGCTTTTAC TAATAAATCT TGGCTTTGTA	8040
TCCGTTTATT TTTTTCATT ACATCAGATT TCGTGGAGCG AGCGGTACCT TTCCCTTGAC	8100
GTTCCTTAA TAAATTTTA TCGACCTCTT CCCGAATCAA ACTTAAGTCT GTGTCCGTCA	8160
ACAATGGACT AACTTTAATA AATGGGTATG TCATTTCTTC AATTGGGAAA ATACTCAATA	8220
CCAAATCGGG CTTTCTGCT TGAATCACTT CAGTTGCATT TAAAACAGAA GCAAATCCTG	8280
CGAGTTCAAC GTTTGGTACT TCCTCTAAAA TTTTGTTC AATTAAACTG GTCACACCTA	8340
AGCCCGTCGA GCAACATAC ACAATCCTAA CAACTTTAGC TTCCTGCTGT TGTCGTTTCT	8400
GAGCTGCCAA AAAATGCAAC GCAATATATG CGCAGAAAGA ATCGTTGACT AACAAAACAG	8460
AACGTTGAAT TAATTGTTTA CAGGCACTTT GAATAGCTAA AAACAACGCT GGGTGTTTTC	8520
CTTTGATTTT TTCAGAAAAC GGATTGTATT CATTAATAAA TAAATGTTTC TTTGTAAAAC	8580
GTAAGGACAG GTGTGCAAC AAGTTCGTAA ATAATTGGCG ATCATTTACA AAAGGATAAT	8640
TCAGCTCTTT TGAAACGCTT TCAATCACTT GCTTTGTAA TTGAACCATA TCTTGCCGTT	8700
GATGGGTTCC GAATAATCA CTATAATAT AAAAATATTC ATCCGCTAAT AAAGGCAAGT	8760
CATACCGTTC AAACACCTCT TGTAAGTA AAAGTGAT TC GTT	8805

## (2) INFORMATION FOR SEQ ID NO: 136:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2738 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

TATTTTTTGC TTAATATAAT TAATGACATC TTTCTTTTGT TCTTTAGCCT CTGATATATA	60
TTCCTTTAGT TCTTTATCCG TAATTGTTGA ACTGTTAATT CCTTGTTTCAT ACAATTGTCTG	120
TCTAGTTTTT GAAACATCAT CTGTTGGATT CAATTCTCTA TTCTTTAAAT TCTCGGTTGT	180

TTCAATTGTT ATAGATGATG TAGAAGTTGG TGTGTGTGCC GGATTACTAC TATCTTTTTT	240
AGAAGACTGG GTAGTTTGCT GCTTAGTTGA TACTTCAGTA GATGAAGCGT CGCTATTTTT	300
TCCCCAGTA TTATTACTAC ATGCCGTTAA AAATAAGGCT AACAATAGTA CAATAAGTTT	360
TTTTTTCATT ATATCCTCCT ATAAAATAAC GAAACATGGT CGGCCTTCTG TCCAATCAAT	420
TGGATCATTT GACGGGGTAT TCCAAAGTGC TTGTGTTGAC ACCCAACCTG AGTAAACTG	480
GCGACCATAC GGATCATAGG TTAAAGTGTA ACCATTTTTG TACTCATGAA CGACAATAGC	540
ATGTCCAATC CCAACATTTT GATATAAAAT AACGGGATAT CCTTCAGAAA GATAGGCGTT	600
TAATGTAGCT ACATCATTA TAACCTTTGAT CGAACGGCCA TATGATTGAG CTGTTATCAC	660
TAAATCAGTT GCAGAAGCGC CAAAGTATCT TTGATTTAAC CCGCCATAAG ATGCCATTCT	720
TGTTGCAACA CTGACAGGAT TTACATTATA ACCGTAAGAG CCTCTTAAAA TCATAGATAA	780
ACTTGTTGGA ACACATCCTG cTGCGCCAAT GGTATAGTTT CCGATATAAG TATTAATCCA	840
TTGTGGGTCA CGTTGAGAAT AATAATAAGG TTTTTCTTT AATCTTCTG CACGTAATCt	900
GTTTGCTATC TCTGCTGGTG TGcGAGCTTC ATCGCCAAT AAAAAGGmTT TTCCgATcCT	960
TTAGGAGCAG TTTCrCCyTT TCTAACTAyT TTGATTTGrA TrGCTTCyAA ATGTTTAGAk	1020
GCAGCTTGAC TTCCAGCTGr GGCrCCATTA GCTGCCCAAT TyAACCAACC TTTATCTTGG	1080
ATATGGACAC GGTAATAGAC ATCATAATAT CTTGAAACAT CACCGmCAG TTTAATTTTA	1140
ATGGCTTCTA ATTGTAAATT TTTTCCAGTT GTGCCAGAyA ATGCCCCATT rGCTTTATAA	1200
kCTTGCCATC CGATATTTTG GATGTGCGTA GAGTATTCTA TrTTTCCAGC TAGAGCAGCA	1260
TCACTCAGGT TGATTTTGAT TGCTTCCAGT TGTAATTTT TCCCACTTGT ACCAGCTATT	1320
TCrCCATTCT TGACAACCCC TTGCCAACCA ATATTTTGCA CGTGTGTTTG ATArTTGACA	1380
kTTGGTTTGA TTTcATCTGG ACgTTTTGCT TCATTTCcAA CTAAAAAGGA TTTTCCaTTT	1440
CCTTTAGGAG CAGTTTCGCC CTTTCTAACT ACTTTGATTT GGATGGCTTC TAAATGTTTA	1500
GATGCAGCTT GACTTCCAGC TGggGCGCCA TTAGCTGCC AATTTAACCA ACCTTTATCT	1560
TGGATATGGA CACGGTAATA GACATCaTAA TATsTkGAAA CATCACCGGA CAGTTTAATT	1620
TTAATGGCTT CTAATTGTAA ATTTTTTCCA GTTGTGCCAG ATAATGCCCC ATTGGCTTTA	1680
TAATCTTGCC ATcCGATATT TTGGATGTGC GTAGAGTATT CTATGTTTCC AGCTAGAGCA	1740
GCATCACTCa GGTGATTTT GATTGCTTCC AGTTGTAAAT TTTCCCACT TGTACCAGCT	1800
ATTTGCCCAT tCTTGACAAC CCCTTGCCAA CCAATATTTT GCACGTGTGT TTGATAAwTG	1860
ACAGkTGGTT tGATtTCATC TGGACGTTTT GCTTCATTTT CAATTAAAAA GGCTTTTCCA	1920
TTTCCTTCAG GTGCCGcTTC GCCCTTCTTC ACTAATTTAA TTTGAATGGC TTCTAATCTT	1980
TTAGCGGCGC TTTGACTAcT GCGGACTCAC TGTTTTTTGC CCAATTCAAC CAACCTTTAT	2040
CTTCTATATG CACACGATAA TAAACATCAT ATGCGTTAGc tATTTCTCCA GTTAAGCGAA	2100
TTTGAATCGC TTCGAGCCGC AAAGACTTTC CGCTTGTAAC AGACAATTGA TTATCTTTCA	2160

CATATCCTTG CCACCCAATT TCTTGACAT GTGTTCTATA TTCTACAGAT CCAGcCAAAT	2220
CTGAATTTGA AATATTCATT TTAATTCCTT CAAGACGTAA AGAGCGTCTG CTA <del>g</del> kcCCAG	2280
AAATTTCCCc GTTTTTGaCA AcCCCTtGGC cAwCCGATGT CTTGGGATAT GTGtCtGATA	2340
ATTAATtCAG gAcTTTTyCT cTGctTTGCC gGTGCTATAG GTGTGtCCTC ATTACTTkGT	2400
GATGTTyCCA AAGTAGTTtC AGTAcGTGCG TACTATyTTC TGTATCATTa ACAGTTGTTA	2460
CTTCTGATGT AGAAACTTCA GTGGTGAAT CGGAGGTGA GGAAGTGGAC GTTGTtTCTG	2520
TTTCCTCTTG AGAATTTATG ATGCTTGGGT TCATTTCTTC AGTGACACCG CTACTCATT	2580
GTCGCTCCGT TAGAGATGTT TCTGAAGCCA CCGTTTCTTC AGTAGTCCC TTAGAGCGTT	2640
CcAAAGCGTA AGCAGAAATA GGGTAACGA GAAATCCCGA CAACAAAACC GCTAACATAG	2700
AACTAGTTAT AACACTTTTT TTCATTAATT TACTCCTT	2738

## (2) INFORMATION FOR SEQ ID NO: 137:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5550 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

GATTGGACGA AAGTAACTAA TAAATCTGTT TTTATAACTG GTGCTTCAGG CATGATTGGG	60
ACCTTTTTTAA TTGATGTGCT TATGTACCAA AATAAAACGA GAAATGCGAA TATTTCGATC	120
TGGGCGATGG GACGAACATT GTCGCGTTTA GAAGAGCGAT TTACTAGTTA TCTAGAAGAT	180
CCGTTATTCC ATATAGTTAT CGGAGATGTC ACAGAAGAAA TTCAGATAGC AGAAGCTTGT	240
GACTATGTGT TGCATTGTGC TAGTAATACA CATCCTAAAT CATATGCCAG TGATCCAATT	300
GGAACGATTA TGACCAATAT TGCTGGTACG CAGCATATTT TAGATTATGC GGTCAAAGCG	360
AATAGTGAAA AAGTGTTATT TTTATCAACT GTTGAAATCT ATGGAGAAAA TCGTGGAGAT	420
CTTGATAAAT TTACGGAAGA CTATTGCGGC TATATTGACT GTAACACGCT ACGGGCCGGT	480
TATCCCGAAG GAAAAAGAGC AAGTGAATCA TTATGTCAAG CCTATATTCA GAAATATGGT	540
ATTGATGTTG TTATTCCAAG AATTAGTCGT ACTTTCGGGC CTACGATGt ACTTTCGGAT	600
AGTAAAGcTT CTCTCAATT TATTATGAAT GCAGTCCATA AAGAAAACAT TGTGTTAAAA	660
AGTGCAGGGA CCCAACTATA TTCTTATGCA TATGTTGCAG ATATTGTATC TGCTCTTCTG	720
TTTTTATTAG TTAAAGGGCA AAAAGGGGAA GCTTACAATG TTTCAAACGA ACATTGCGAT	780
ATTACGTTAC GTACTTTTGC GGAAACATTA GCCAATGTGG CTGGAACAAA AGTGATTCAT	840
GGAGAAGCCA CCGTCCAAGA AAAACAAGGG TTTTCTAAAG CAACAAAAGC ACTACTAGAC	900
AATCAAAAAA TCTACGCATT AGGTTGGCGG CCACTGTACG ACAGTATGGA AGAACCATTA	960
AAACATACAA TTAAATTTT AGAAGGTAGC GGAAATTGAT ATGCCCAAAA TTAGTATTAT	1020
TGTTCTGTGA TACAATGTAG AAAAATATTT AGAAAAATGT GTACGCTCTA TTTTAGCTCA	1080

AACGTTTACT GACTTTGAAT TAATTCTGGT GGACGATGGC TCTCCAGACA GTTCTGGAGC	1140
AATGTGTGAT CAGTTTGCTG AACAAAGATCA ACGGGTAAAA GTTATCCATA AAGAAAATGG	1200
TGGGCTAAGC GATGCTCGTA ATGCTGGAAT TGAAATAGCA ACAGGTGAGT ATTTAGGTTT	1260
CGTAGATAGT GATGATTACA TTGCAGATGA TATGTATGAA CTATTATATA CAAATATCGT	1320
AAAAGAAGAT GCCGATTTAT CAATCTGTGG TATTTATGAT GTCTATGAGG GGAAAGAGCC	1380
AATTGTAAAA AGTTTAATAC AGGGAACTTT TTCCAGAGAA GAAGCATTAT TGTTAATTTT	1440
ACAAGGAAAT ATTATCTCCG TACATGCTGT AAATAAGTTA TATAAACGAA AACTCTTTGC	1500
GGATTTAAGA TATCCTAAAG GAAAGTATCA TGAAGATTCA TTTATTATAG TAGATTTGTT	1560
GTCTGAATGT CAGAAAGTCT CTATTGATTC AACACAAAAA TATTACTATT ATCATCGAAT	1620
GGGCAGCATT AATACAGAAA CATTTTCAGA TAAACAATTT GAGTTTATAG AAGCTTGGGA	1680
GAAGAATGAA CTTAAATTGA AAGGTAAGGG AGCAGTTATA GAAGAAGCTG CCCATCAAAG	1740
AGTTTGTTTT GCGAATTTTT TAGTATTAGA TAAATTTTG ATTAGTAATG CTCCGAAGAA	1800
AAAAGAAACG AAACAAATTG TTCGCTATCT TCGGGAAAAC TTTATTTTAA TTATGAAAAA	1860
TAAAGTTTTT ACAAAAAGTC GTAAACTATC AATGATTTTA CTTATGTTTG GTCTTCCTTT	1920
TTATAaAATA CCAATTAAAT TGAAAAGAAA ATAGTCAAGT CCAGACTCCT GTGTAAAATG	1980
CTATACAATG TTTTTACCAT TTCTACTTAT CAAAATTGAT GTATTTTCTT GAAGAATAAA	2040
TCCATTATC ATGTAGGTCC ATAAGAACGG CTCCAATTAA GCGATTGGCT GATGTTTGAT	2100
TGGGGAAGAT GCGAATAATC TTTTCTCTTC TGCGTACTTC TTGATTCACT CGTTCAATTA	2160
GATTGGTACT CTTTAGTCGA TTGTGGGAAT TTCCTTGTA GGTATATTGA AAGGCGTCTT	2220
CGAATCCATC ATCCAATGAT GCGCAAgCTT TTGAATATTT TGGTTGATCG ATATAATCAT	2280
GAATCAATCG ATTTTTAGCC TCACGCGCTA AGTTAATATC TGTGAACTTA AAAATTCCTT	2340
TAACAGCTTC TCTGAAAGAT TTTGAATTTT TTTTAGGAAT GGTGGTAAAG ATATTTCTTA	2400
GGAAGTGAAC TTGGCATCTT TGCCAACTTA CGTTGGTGAA GGATTTTCTA ATGGCAGAGA	2460
CTAATCCTTT GTGCGCATCA GAAATAACGA GTTCCGTACC TTGTAAACCG CGTTCCTTTA	2520
GGTATTCAAA AAATGTTGTC CAGGTCTCTT CGCTTTCGCC ACTTTGAATC ATGAAGCCGA	2580
TAATTTACAG GTCGCCATCT TTGGTTATTC CAATCGCTAT ATGACAGCTT TTTGAGAGTA	2640
CTCGATTTTC TTCTCGTACT TTTATATAGA GTACATCGGT CATTAAAGTAA GGATAATTTT	2700
TTTCTGATAA TAAACGATTC TGCCACTCGT TAACCATAGG TTCTAGCTGT TCTGTTAAGC	2760
TAGAAACGAA GGACTTAGAG ACGGATTTAC CACAAAGTTC TTCCACAATT TTTGATACTT	2820
TACGAGTTGA AACGCCTGAT ACATACATTT CCAACATTGA AGCCATGAGG GCTTTTTCGT	2880
TTCGTTGATA ACGTTCAAAC ACTGTGGGTG AAAAATGGCC ATCACGTGTT CTGGGTACTT	2940
TTAATTCTAG CGTGCCTACA CGTGTCTGTA AGcTGCGCTC ATAATAGCCA TTTGTTGAC	3000
TTTGTCGGTT TTCTGTTCTG TCATATTCTT TTGCTTGAAT ATATTCTGTT CGTTGATTTT	3060

CCATTAGTTG ATTAAATACC GTTGTAAAA TATTTT TAGA AACGTCATCC TTTACAGAAT	3120
ATTCAATAAT GCTTTGAATC TCTTCGCTTT TCAGTGTAAG ATGTACTTGG GTCATGTAAA	3180
AGTCCTCCTG GGTATGTTTT TGTCGTAAAA AACATTGTAC CGTAAAAGGA CTGTTATATG	3240
GCCTTTTTAC TTTTACACAA TTATACGGAC TTTATCGACA AACTGATAAA TTTAAAGATT	3300
CAGTAAATAA TTCCATCGAA GATATTATCA ATGAAAAGAG TTTAACTTTA AGTAATCAAG	3360
TAAAAGAATT AGAAAAAGAA TATGAAAGCA AAATACAAGA AGCAGAAACA ATAAAAGATG	3420
AGAGTATATT AAAAGCAGCA ACTCACCCTT ATGAAAAACA GAAACAAAAA ATTGAGGAGT	3480
CTTTTACAGA AGAAATTAAC AAAGAAGTTG AAAAAACGAT CAATACTGTT GTGGAAGAAC	3540
AAATAGAAAA AGTGAAGAA AAGAAAAAGA AGACAACAGA AGATGATGTT CGAGATCATT	3600
TACGAGGATT TGCCCGTACA ATTCCTGCAT TCTTAATGGC ATACGGAGTT GAAGATACAA	3660
GATTAGAAAA TTTTGAGGTT AATATTGATG AAGCCACATT CGAAGATTG ACGAGTATTA	3720
CAATAGAGGA ATTCAAGAGG CTCCGTGATG GATTTGAGTA TCCAGATAAT AATGGGGAAA	3780
CTAAAGTTAT CCCAGGATTA TTCAATGAAG TTGTATTTAA TGCTAGTATC CAGGAATTTT	3840
TGGAGACAAA AAACCGTTTG GCAAATTATT TTGATGATAG CTTGCAAGAG GACATTTTTG	3900
ATTATATCCC ACCACAGAAA ACAAATCAGA TATTTACACC AAGACGTGTT GTAAAATCGA	3960
TGGTAGATTC CTTAGAAAAA AGAATTCCAG GAATTTTTTC GGATAAAACA AGGAAATTTG	4020
CAGACTTATA TGTTAAAAGC GGTATGTACA TTACAGAAAT AGTCAAGAGA TTAAATCATG	4080
GATTAAAAGA TCAGATTCCA AATCAACAGG AACGTATTAA ATGGATTCTT GAAAATCAAG	4140
TTTATGCATG TGCTCCAAGT AATATTATTT ATAATATTGC CAAGAACTAC ATATTGGGAG	4200
AAATGGAGCA TGTTTCTAAT ACAAATATTA TTGAAGTTGA TAACTGGCT TTGGCAAAGG	4260
AAGCAAAGTT ATCAGAATGT ATTTATAAGG AATTCGGAGA TGTTAAGATG AAATTTGATG	4320
TAATTATAGG AAATCCTCCA TATCAAGAAG AAAATACTGA AAGGAATAGA GATGACGCTA	4380
TCTATCACTT ATTCTTAGAT GAGGCGTACA AGATCGCAGA GAAGGTTGTT ATGATTACGC	4440
CAGCAAGATT TTTATTTAAT GTTGGATCGA CTCCAAAAAT TTGGAATGAA AAAATGCTAA	4500
ATGATAAACA CTTATCTGTT GAAATGtATG AACGGAAAAG TGaAAAAATC TTTCTAATA	4560
CaTCgATTAT TGGTGAATC GCAGTCACTT ATCgTGATGC CTCAAAAATC CTAGGACCTA	4620
TTGGAACTTT TACACCGTAT CAAGAATTAA ATTCAATATT GAGTAAAGTA AAAGAAATTA	4680
ACCAAGAAAG TAACTTTAGT TCACTGATGT TCGTACAAAA TAAATTTAAT TTAGATAAGC	4740
TATATCGTGA TTTTCCAGAT TTTGAAACAC GTCTTGAGG AAATGGACGT GAACGCCGTT	4800
TAACATCTTC AATATTTAAT GTATTACCTG AAATTTTTTC AGATAAAAAA ACAAAGAAC	4860
ATAATCTGGC AATTTACGGT AGAGAAAaTA ATAAGCGAAC aTATAAATAT GTCTCTAAAG	4920
AATATATTGA GGAACATCCA AATATGTATA AATGGAAGGT TTTTGTTCCTA GCAGCAAATG	4980
GTGCTTCGGG TACAATAGGA GATAGTCCAG CAAGAATTAT TAGTTTCCCG GTCTTAGGTG	5040

AACCTAATTC TGGTCATACC CAAACCTTTA TAAGTTTAGG AGCTTTTGAT ACGAAGTTTG	5100
AAGGTGAAGC GTTGTTAAAA TACATCAAGT CAAAATTCGC CCGTGCGATG CTAGGTGTAT	5160
TAAAAGTAAC GCAAAATAAT AAGACTCAAG AAACATGGTC AAAAGTGCCT GTATTAGATT	5220
TCACAAAAGA TTCTGATATC GATTGGGAGA AACCCTTGCC TGAAATTGAT CAGAGTATAT	5280
ATAAAAAATT TAGTTTGAAT CAAAATGAGA TTGATTTTAT AGAAAACCAT ATAACGTACC	5340
TTGATGAAGA AAGTTTAAAT AATTTATAAG TCAAGATAGA TACTTGTGAT TAAAAAGTA	5400
ATTGTATTCA ACTGGCATCT ATTTTGGTT CAAATTGGTT CAAGTAATAT CAATATCAAT	5460
CTATTTCCAT ACATTTTAT CTATATAACA TTTCCAAAA CGTTGATATA AATGGATTAG	5520
GATGTATTCA ACACACGTCG TATTTTATA	5550

## (2) INFORMATION FOR SEQ ID NO: 138:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28690 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

AGAAGAAAAT AnCATCAAAT ACGCTTTCTC TAGATGGGTT AAAATGCTTG GAGTACACAT	60
TACGATACGA TTGTTGCTTC CATTGTGATT CCTTTTGAAG AAAAAGATGG CTnTGTGACA	120
GAACTATCTG CAGGCCAAGC CTTTAACCGC ATGATGGGAG ATCCTGACAA AGAGGTTCGT	180
CAACGTTTGT TTACAGCCTG GGAGAAAGCT TGGAGTGGCA AAGCTTCAAT TTTAGCAGAC	240
ACCCTAAATC ACTTGATGG GTTCCGTTTA TCTGATTACA AATTGCATGG CACTACAGAT	300
TATTTAGAAA ATCCCTTAA CTATAATCGA ATGAGTCAAG AAACATTAAC GGTATGTGG	360
GaAACCATCC AAAAAATAA ACAACCATTG GTCGACTTTC TAACGCGCAA AGcGCAAtTA	420
TTTGGCAAAG AAAAAATGGA TTGGCAAGAC CAAGATGCGC CAATTATTTT AGGTGATTTA	480
GAAGAAAAAA CGTATTCCTT TGATCAAGCC GCTGCCTTTA TTTTAGAAAA TTTCCGTAAA	540
TTCACTCCAA AAATGGCTGA TTTTGCCAG ATGGCTTTTG AAAAAAGTTG GATTGAAGCT	600
GAAGATCGTC CAGGCAAACG TCCTGGGGGT TATTGTACAG AATTACCTGA AACACAAGAG	660
TCGCGTATTT TCATGACCTT TGGCAATTCA ATCAACGAAG TAGCTACTTT AGCCCATGAA	720
TTAGGTCACG CTTTCCATAG TAGCGTGATG TGGGATTAC CTTCAATTGAA TCGAGAGTAT	780
GCCATGAACG TAGCAGAAAC GGCTAGTACC TTTGCAGAAT TGATTGTTGC TGACGCCACC	840
TTAAAAGAGG CCAAATCAGA TGTGAAAAG ATTAACCTTT TAGACACAAA AATGCaAAAT	900
GCCaTTGcCa TGTTTATGAA CATTCaTGCG CgTtTTATCT TTGAAAATAA TTTTtATCaA	960
GCGCGCCAAA AAGGTTTAGT AGCGGAAGAA GATATTACTG AATTAATGTT AGCGGCGCAA	1020
AAAGAAAGCT ACAAAGATAG TCTTGGCAGT TATCATCCTC ATTTCTGGGC GAGCAAATTG	1080



CACTTCTTTA TTGATGATGT GCCTTTCTAT AATTTCCTAT ATACTTTTGG CTATCTATTC	1140
AGCATGGGTA TCTATGCGTA TGCCAATCAA CAAGGCAGTT CTTTGAAGA CCAATACATT	1200
GCGTTATTAA GAGACACTGC TTCGATGACA AGCGAAGAAT TAGCGAAAAA ACATTTAAAT	1260
GTTGACTTAA CAAAACCTGA TTTCTGGCAA GCCGGCATTG ATCAAGTCCT AAAAGATGTG	1320
GAACAATTCA TGACTTTAAC TGAAACTAT GTAAACTAAG CAAAAAGAA GCAGTAACCA	1380
ACATCAGGTT ACTGCTTCTC TTTTTCCTCA ATAGTTTCGA TACGCAAGGT AGCCACAAG	1440
CATTAGTCCA AACGTAACAA ATTCAACTAC TGGCATGGCT AACCAACAC CGGTTAATCC	1500
TAGTAATTTT ACCAAAAATA ATAAACTGG CACAATTAAA AAGAACCAC GCAAAGAGGA	1560
AATCATTAGC GCCATTTTCG GTTTACGAAC TGCCGAAAAG TAATAAATCC CCATAAATT	1620
CAATCCTGTA AACAAGAAGG AAATCGCATA TAAGCGGATG GCTGTCCTAG CAATTTGCGC	1680
CAACTGGTCG TTCCTTCGC TATTAAAGAC CTCAATAATT TGTCCCGTAA ACACTTGACT	1740
GATTAAAAAG AACAGGAAGC CGGCGACCAA GTACGTAATA AAGCTAAGTT TAACAAATTT	1800
ACGCAACACC TTCGTTTCCC CTAAACCATA GTATCGGCTG AGTAATGGTT GCGCCCTTG	1860
CCCAATCCCT GTAAAAATAG CAATCACAAT AATATTGAGA TTGGCAACAA TCGCATAAGC	1920
TGAAATCGCC ACATGACCAA CTAAGTTTAA CAAAACAATA TTAAATAAAA ACATGACAAA	1980
GGCCGAAGAA AATTCATTTA AAAAGGAGGA AAAACCAATC GACATAATTT CTCGAACATC	2040
CTGAAATTTT ACTTTAACT TTTCAAACG TAAACTCGT TCGGCGACC ACTTATGCAA	2100
CGTAATCAGG ATTAAGCTTG TTAACGGTGA CATCACCGTT GCAATGGCCG CACCTGCCAT	2160
ACCCACCCA AATTGATAAA TAAAGACATA ATCCAAAATA ATGTTGACAA TTCCCCACT	2220
GACAAAAGCA ATCGTTGCCA AACGAGGATT ATGGTCATTT CTAAAAAAG TAATCAGGCA	2280
ATTGTTTAAA ATAAATAATA ATGAACAGCT AAGGTAAATC ATATAATATT CCTTGCCAA	2340
ACCAAACAAG ACACCGGTCC CCTGAAGCCC CCGAATAATT TGATCGGAAA AAATCAAGCT	2400
CACTAAAGTA AAGAGACTCC CTATAACGAA AACTAACCCA ATCGTTAAAC TAAAATACTG	2460
ATTGGCTTTT TTGATTTCTT TTTGCGCAAC CGTGGTCGAA AACAACGTAG CGCCGCCAAC	2520
GCCAAACATC CAACCTAATC CATTAAATAT ATTAACGAGT GGTAAGACAA TGTTAAGCCC	2580
CGCCAACGCC TCGGTGCCGA TCCCATTTCG AATGAAAAAT GTGTCCGCCA AAATAAAAAT	2640
TGATTGCGCC GCCATACTTA AAATTCCTAA CAGAACATTT TTATAAAACA ACGCAGTAAA	2700
ATGGTTTTCG TGTGTTTGCT CCATACCCAA ATCTCTTCTT AACAAAAAAA GGAATCAATC	2760
TTTTTGCTTC AAGGTTACGC AAAAATCAGA ATCCTTCAAT TATTCAACCC AGCGATTAAA	2820
TAATCTGTCT ATTTATTGAT ACTTCTAGGG TAGCATTCTC TCTTCGAAAA AACCAACATC	2880
CTTTTATTTT TCTCATTTTA AACTTTCTTA AGCAGTTGCA GTTGTTTTTG AACCATTGCG	2940
CAATCTTCTT TCACGACTTC TTTTAAACGA TAATTATAAA AGACAGCCGC TGGATGGAAG	3000
AGTGGAATGA TTTGATAGGT TTTCTGCGTC CAATGATAGT TTTGTGATTG CTCATCAAAT	3060

GTTTGAATGG GCGACTGAAT GACCGTTCCG TGCCTGACAG TGATTGTTGG CTTTGGTCCT	3120
AAAAGTCGTT GCAACCCAAT ATTCCCCAGT GTCAAAATTA GTTCAGGTTG GCAAACAGCG	3180
ATTTCCCAAT CAAGCAAAGG TGCAAAAATT TTGATTCTCT CTTTGGTTGG CGTTCGATTA	3240
GGAAATTTGG TTACTCTTTC TCCCGTTTTT GCAATCTGCT TCGTTTTTAC AGAAAAAGGC	3300
CGACTCCTAA CCGCACTGGT GATATACACC GTTTCACGAC TTAAGCCAAC TTGTTGTAAT	3360
TGTTCCATCA ACAATTGACC TGAACGCCCT ATAAATGGTA CGTTTTCTTT AATTCTTCT	3420
TTTCCAGGGG CTTCCCAAT CAGCATCACT TTTGGCTGTT GCGGACCTTG ACCTCCACA	3480
AAACCTTCTA ATGGTAAGTC TTTGCTTTA GCTAAACTG CTTTTTTCAG ATGTTTCAGGA	3540
TAGTGCATTT ACCATCACGC CTCTCTTCTG GTGTTCTTCC TTCGCTATAT TGATTATAAA	3600
TAAGGCCCAA AAGAACTCA AAAAAAGGAG CCTATTCATT TAAGCGCTGG TTTTCTACA	3660
CACAAAGCTT TGTCTTTAGC TAAATAAAAA GGCCGTAAGT AACCAACTAA ATTAGACATT	3720
GGTAAACCT CCGATGGTAA ATCGAGCTTT AATGTCTCCT TTAAATAGGC TTTGCGTGTC	3780
GCAATCCGTT GCCACATGTC TGGATAAGTT TGCTGAATCT CTTTTTGCAG TGTTTCATCA	3840
GCCAAAGCCA CACATTCTTC TGCACTCACC CCTGTATAAC CAGGTACAGA AGGAATAATA	3900
TCCACTTGAA AGAGCATGCC ACTTTCCAAG CGAATGGCTG AATCTGGATA AATAGGCGAC	3960
GACATCCATT CATCATCCGA AACTAAATGC CCAGGATTCA AATGCCAGTG ATAGATTCT	4020
TTTGCTAACT GTTCTTCAAT CGCTTGGTAC ATTTCTCGCC CTAACAGTCC GATTCTGAATG	4080
GTTTCAAGCC AGTGAACGAC AGCTTGAAAA TAGGGTTTTG ctACGCGCTC TAAATAATCA	4140
CGTTGCGCTT CTGGCAATTG CTGCTCGTTT TCAATAACAA AACCTGTTCTG GCTAGACAAG	4200
CCGCCTTTGT AGCCTGTAGT TAATGATAAG GCATCACCTA ACTGAATTTC TTTGGCAGTC	4260
GGATACATGT TTGCATATTC AAAACGTTGA CCAGTTGCCG CAATGGTTAC TACCGTCGGC	4320
GTTTGACCTT CATCATTTAA CAAAGCCCCA ATGTCCGTTT CACGCTGTCC TATTTCAATT	4380
GCATTTAAGG CCTTTAACAT ACTGCGAGAA GCCAAATTCG CCCCATATTC ATAGTGCGCC	4440
AATTCGTTAG CATTATTTGT TGCGCGAGCC CCCTTGCTC CAATCAATAA GTGGGTCCCA	4500
TTTAGTAAAC GAGCTTCTTT CGGCAGCGCT TTTTTCAGTG CTTCGACAAT AAAATGTGGC	4560
ACATCAAACA TAGTAGCAGG TTCTTGCTGT TGTGTCGTAA ACATTTTCCA CCCAACGATA	4620
CCTATTTTTT GAGCAGTTTC ATCGAGAAGC GTTTCAAAA TCTGCGACAA CTTTTGTTCC	4680
CCAGCCATTG GTTGATTAGG TAGTGAAAA GCTAGATAAT GAATTAAATC GGCTGAGATG	4740
CGGGCATGTT GGCACAGTTT TAAATTTTCA TTACCTAAAA TAAGTGTCGC CGCACCATCC	4800
TTATTTAACA CAAGAAGTCC TTCTTCAAAC CTCGGAATAA AGCCAGTTAA GTATTCAAAA	4860
TTACTGCCAT GCTCTTTGTC TGCATAGAAG ACCAATTGAT CGAATTTTTC TGTTTCCATT	4920
CGTTGTAATA CTTTCTGCTT GCGCGCCATC ATCGTTTCAT TCGTCAAATA CGTTGGAAAA	4980
ACATTCTCAA AACAGTAGG TTCAGCAATA TGTGTATAAT TAATTTTTTCG CTTTATTATT	5040

ATCCTCCATT	ATTCATCTGC	GACAATCAAC	GTTACGCCTG	CTTGCCTAA	CTGCTCACAG	5100
AATTTTTCAT	CAGGTCGTGT	ATCTGTCACT	AAACAATCAA	TTTCTGTTAA	TTGCAAATT	5160
TTAAATAAAC	CCGTTTGTG	TAATTTCTGT	TGGTCTGCTA	AGGCAATGAC	TTGTTTGCCT	5220
TGTTGAATCA	TTGCGCGCTT	GACTTGACCT	TCTTCTTCTG	TGTGAATAAA	TAATCCTTGT	5280
TCAGATAACG	CAAACACACC	TAATAATGTA	TAGTCGGGGG	CAAATTCCCC	AACCTTAGCG	5340
ACTGTTTCTG	CCCCAAACAA	AAAGAGTTGT	TGTTTATGCA	GTTGCCCAGG	TAATACTGTA	5400
ATTTTGTCT	GATCCATTG	CGCTAGAAGT	AGCGCATGAC	TGAGAGAGTT	CGTTATCGCG	5460
TGAATCGGAC	TATTTTCCAT	CGCTCGAATA	GCTGCTTCCA	CCGTGGTTGA	TGAATCAAAA	5520
AAAATGGACG	CTTGTTTGGG	AATATTGACC	CAAACTTTTT	GGGCAATCGC	TTGTTTTTTC	5580
TCTGGATGCT	TCACTATTCG	TTGAGCATAA	TCGTAAATAG	TTGCTTCAGA	CTGCACTGAG	5640
ATAGCGCCAC	CGTGCCTTCT	CTTTAAATGT	CCTTCCTTTT	CTAAACGAAT	TAAATCTCGC	5700
CGAATAGTAT	CTTTTGAAAC	GGAGACTTGC	TnCCGTAACG	CTTCCATGGA	AATACTTTTT	5760
TGTTCAATCA	CTATTTTTTG	AATCTTGGCT	AAACGTTCTT	CAATCAACAT	GCTTCCTCAC	5820
CTCTTCTAAG	TAAAGATAGC	ACCTTTTGCA	TCATTTTGCA	AACAAATGCA	AAACAATGCA	5880
AAAAAAGACA	AAAAAACTAG	TCCTTCATTT	TCTAACGAAG	GGCTAGTTTG	AGAGTGGTTT	5940
TAATAGGTAA	ATTGAATAAT	CATGCCGCTA	GGATCTTCTA	ACGCAAGTAA	ATGATCTTTG	6000
GCAATAAAAT	GTAACTTTGC	TTCTTCCGCA	TTGGCTTGGA	CTATCTCTAA	ATCTGCCGCA	6060
GTTGGTAAAC	GGAAGGTGTA	ATTTTCTAAG	CCAAATTGAT	TTTCTTGAAT	CGTTGGCAGC	6120
TGTTTGCCGC	TCCATGTATT	TGTCCCAATA	TGATGATGGT	ACTGACCCGC	TGCAAAAAAC	6180
TTGCTTGTG	GACCAAAGTT	AGACTTTAAG	CCAAACCCTA	AGTTTTTATA	GAACGTTTCA	6240
GTGGCATCTA	AATCGGCCAC	TTTCAAATGA	ATGTGCCCAA	TTTAGAACT	TCGGAAATTC	6300
CTAGCCAGTG	ACCGTCAGCT	TCTGCAGCTA	AACTATCACC	GTCTAATTCT	TCTGTCACAC	6360
CGATAATTTT	ACCATTGGA	CGAATATCCC	AGACTTCTTC	TGGTTTATCC	CAATAAATTT	6420
CGATGCCATT	ATTTTCTGGA	TCACTTAAGT	AAAGGGCTTC	ACTATAGCCA	TGATCTGctG	6480
CGCCAATTTG	CACATTTTCC	TGCAAAAGCC	ACAGCAACAT	ATTGCCTAAA	TCTTTGCGGG	6540
TTGGTAGCAG	GAAAGCAATA	TGGTACAAGC	CAGTGGTACC	CTGATTTTCT	TCCCCAGCTG	6600
TTAGTTCTTT	TAAGACAAGA	ATTGCATCGT	TGGTGCTTG	TGGCGCTAGC	CAAGCACGTC	6660
CCTCTCCTTG	TTTTAAAAGT	TTTAAGCCAA	TTGTTTTCTG	ATAAAAGTTT	ACCATTTTAT	6720
CTAAATCAGC	CACTTTCAAA	GTGACTTGTT	GAATCCCAAT	ATTACTTGGT	AATGCTTGAA	6780
CCATTTGCTT	TCACTCCTTT	CCTCTTCGTT	TTACTAAAGC	AATGATTTAG	TTGAATTCTT	6840
TAAACGAGC	ACTATCTTCC	ATGTATTCTT	TTTCGcCAGc	tGGCGCTTCA	ATGGAACCAA	6900
TGACCATTG	TGCACGTAAA	TTCCAGTTTG	CAGGAATTGA	CCATTCTGCA	GCAACCGCTT	6960
CGTCAATAAC	TGGATTGTAG	TGTTGTAGGT	TTGCGCCGAT	ATTTTCTTGC	GCTAAGGCAG	7020

TCCAAACGTT	CGCTTGTT	AAACCAGTTG	CTTGTTTCAGA	CCAAACAGGG	AAGTTATCCG	7080
CATATAATGG	GAAGCTATCT	TGTAAGTTTT	TCACGATATC	TGTATCTTCA	AAGAATAAAA	7140
TTGTGCCCCG	ACCAGCTGCA	AAACCTTGTA	ATTTAGCTTG	TGTGTTTGGA	AATGCTTCTG	7200
CTGGAGTTAA	TGGTTTTAAG	GCGTCTTCAG	TCATTGCCCA	TAATTTTTTG	TGAGCATCGC	7260
CGAATAAGAA	CACAACACGT	TGTGTTTGAG	AGTTAAAAGC	TGTTGGACTT	TCTCTAACAA	7320
CTTCTTTTAC	TAAAGCAGTT	ACTTCTTCTT	TTGAAATTGG	TAAGTTTTCT	CCTAAAGCAT	7380
AGATTGAACG	ACGGTTTTTC	ATCATTTCAG	TAAATTTtGA	CATAATAtGT	TCaTCCTTtC	7440
GAATTtCCTC	AGTAAATyCA	TTaATTnTGT	KTACATGCTT	ACTATACTAC	CTTACTAAAA	7500
GTAAGTAAAy	CTTTTTGTCT	TACTTTTTGT	AAGTTTTTTT	TATTTTTTAG	TGAAAAaCCT	7560
TATTCTACCA	AAGAAAAAAC	AACCAGCTTT	TTTCTAAGAA	AAAGAAAAAT	CTGGTTGTCG	7620
ATTAATAAAT	GTAGAAGTCA	ATTGAATCGT	TTTCGGCCGT	TTTTCTTGAA	ATAAATGTGT	7680
GGCTTTTTCC	CCTAATTCCA	TTAATGGATG	ACAGATTGTT	GAAAAATTTA	AAATTTCACT	7740
TAATAAAGTT	TGTTCCCTGC	CCACTAAAAA	GAGTGGTCTT	TCAATAACGG	CTTGAATCCC	7800
TGCGGCAATT	TCATCACCGT	TGGTCACAAC	CGCTTGGAGC	GTTGGATCTT	GCCCACAGTG	7860
ACGTCCTACT	TCACACCCAT	CTTCATAGGT	TAAACAGTCC	TTAAATAACC	GATTTTCAGG	7920
AATCGCTCCA	TAAATTTCTT	GATATGCGTC	AACTAAATAG	CGTGCGCTTG	TACTTTTTTC	7980
TGGTGACGG	TTAATCGTAA	TTGCCACTTC	ACTACATCCT	TGTTGTTcAA	TGCTnGGAAG	8040
ACCCGCCGAT	AAACGGTTAA	ACGATCGATA	TAGATACTAG	CATAACTAGG	GTCTTCTATT	8100
TTnTCACAAA	AAACAACCTG	CTCTTTATAT	GGAGCAAACa	CaCTTAArGG	ATTTGCCTTT	8160
GaCAAAACAA	TCATGCCaTC	aAACGCACCC	nGTGGCCcAT	TTTTTCCmAG	tAACTTTTTT	8220
nCAACTTCTT	TATCaTAATC	CgTTGGTAAT	AAACTGACCA	TCATCCCTTG	TTGAAATGCT	8280
TGGTGGATAA	TCCCACTTAA	AATAGCATCA	AAATAAGGTT	GATTCGTATA	AGGAACCATG	8340
ACACCTAATG	TATTACTTTT	CCCAAAGCTA	AGATTTCTCG	CTAATATATk	GGGCTTATAG	8400
CCCaACTCAT	CGATAATTGC	TTGGATTyTC	AGTCGTTTTT	CTTcAGAAAC	ATAAGGATGa	8460
TGATTAATCA	CACGCGAAAC	AGTCGTAAC	GAAGATCCTG	AAAGGCGCGC	AATTTCTCGA	8520
ATTGTTGTCA	TACTTTCACC	TCATTACTTT	ACAAAATTGA	GTTAAATAAG	GCACAATCCC	8580
TGTCCCGTAA	TCAACAAAAC	GATAAAAGGT	CTCCTGCCAA	GCAAAAGAAA	TTTCCTGAAT	8640
ATCCTCCTGC	TGTGTCAAGT	GcCATCCGTT	ATCTGATGGT	nACGGrATTT	CTTGATGCCA	8700
AATCTTATAG	TATTTTTTGG	GACTATTTAA	CTGCAGACTT	TTGGCCGTCG	TAGTCACsGr	8760
GACAAAATAG	TAATCCATAC	CAAGTCTATC	GCTTATCTCC	ACTATAACAT	TCATTTTCAT	8820
GCCCTCCAAT	ATTTTGATAC	ACCTAGCTTA	CCTTATGAAA	CGCGTACCAG	AGCAAGAGGA	8880
AAAATAAAAA	AAGTGAGACC	ATGTTATACT	AGATGTATCA	TATAAAAGCA	AAGGATGAGT	8940
GCAATGGAAA	CTGTTACAGT	AGCAACAAAA	AGCGACTTAC	TAACAAATTT	GAACAAGCAA	9000

821

ACAGATGAAA TCATTATCTC GGGGGCACTT GCACAAGAAA TTTACCACTT AAAAAACAA	9060
CAATTACCG AAACGGAAGA AATGGGCTTT TTAGTGGGCA GCCAAGGTGC TGGCTCATTT	9120
TTACCTATT TAATCGAGAA GTTATCCTAT ATAAATTCGG CGCTATCGCC AGAAGAACAA	9180
AAATTGARC ATGCGATTTT GCGTTTGTAT GTTATTAAAA AATTAACCTC AGAACTACT	9240
TTATTACGTC TAAACAGTT AGATTATTAA AATTTTATTT AATCAAGATA GACCAATTCC	9300
CCGCGTACAA CGAGTCCCTT TCATTAATCA CTTGATATAA TTACCTTTAT TCAAGTCCC	9360
ACCATTTGCA TTATCGATTA ACCAATGGTA TATTAGAAGT AACAGATAAG TCATTTTATT	9420
TGAAAAAGTG TTTGGGAAAT GTTTGGGAG GAATGTTTAT GTGTCAACCG CAGGAAAAA	9480
AGAAAGAAAT TTCGTGTGCT AAGTTACGTT ATCGTTTGA AGTACCAATG ATTATCATTG	9540
GTTTTGTTTT ATTGTTTTTC GCATTGGGGT TAGTTGCCTT ATTGATATTG AGTGATTGTC	9600
CGATTCTGCA CTGGTTATTA ATTGTCTTAA TTGCACCAAT TGCGCCAGTC ATTGCTTTTT	9660
TTACCAATTCG TTATATGTAC ATGGATACTG CTTCTAACGG AGTGGaAATT ACAGAAAATC	9720
AGTACCAGA AATTTATTCC TTGTATATTG ATTTGGCAA AGAAATGGGC TTTGGCACTA	9780
AAAAAATTAA GGATGCCACG TCTTTATTTA ATCAACGGGA ATGGTGTCTT AAACGCTTTT	9840
GCGGCAAAAT GCTCTTTACA CAGACGCTAT GTTGTCAATC ACAGTGATTT GTTGATATC	9900
GCGTACAAAT CAGGCGATTT TTCGCTGATT CGCTTTATTT TGGCACACGA GCTAGGACAC	9960
CATAATGTG GGCACACTAA TTTATGGCGA CTAATGTTAT CCATTATTTT GAAACCAGTT	10020
GCGTTAGATA AAAGTTTTAC AAGAACGCAA GAATATACCG CTGATCGCGC AGGTTTATAC	10080
TACGCCGAG AAGGTGCTTT AAGTATGATT TATTTATTTT CAGGAAAATA CATGGGCAGT	10140
CGTGTGAAT TAGAAGAATA CTTTCACAGT ATTGATTAC ATGATGATAC CATCTGGCTT	10200
AAATTGAGTA ACTTCTTATC CGATCATCCT GTTGGCTTCC GTCGAATGCA AACATTGAAA	10260
AAAGCCPAG AACTTGCAA TTGGGACGTT CATGGGAAAT TTTCTAGCA AGAGCCAACC	10320
ATCATTAAAT AAAGTGCTG GGACAAAAGA TTCTTTAACT TTTGTCCAG CACGTTTATT	10380
AATCCACCAA TAAATGAACA AATTCTTCTT TGGTGATATT TACAAAGTAG TCTTCAAAGT	10440
TAAACGCTGC TAATGCTTGC CAAATTGATG AATACTGATA GTCCACACCT AATAAAGCTG	10500
CAACGATTTC CGTTGTTTCT TTTTGGCAA AATAATCACC ATAAATCGTT AACTCAATTA	10560
TTTTACCTTT TTCAACAGTC AAACGTGCAT CCACGATGCC GCCTTTAAAC TTTCTTCTT	10620
TTTTGATCGT AAATTTAGGC GCTTCACCAA AAATCCAAGC TTCATTGCCA TACACTCAG	10680
CGACCAATTT CCTAATTTCC TGTGATCTG CTTTGGTTAG GACATATTCT TTCTCCGCAA	10740
TTTCAGTTAG AGATTCGACA TCAAAGAGCT CCATCAGTAA ACGATTGCGA AATTCCTCAA	10800
TGGTTAATG TTGATATTTT TCGTCTAAGT ACGGCCGTAA ATTGGTTACG CGTCCCCGA	10860
CAGATTTtGT CCCTTTTGAT TCAATTTTTT TCTTAGAAAC AGTTAAAACA CGTTGAACTT	10920
CTGCCAATC AACATCATAC ATCAAGGTAC CATGAGTATA CATCTCCCC TTTTTGTAT	10980

822

ACATGGCAT= GCCAGAAAAT TTTTCCCAT CAATTAATAA ATCATTGCGA CCGCTAATCT	11040
CGGCCCCCGT TGGGCCCATT TTATGTATG CTTCAATAAT TGGTTTAGTG AACGCTTTAA	11100
AATCCCCAA AGCTTGATGT CCTTCCTGA CAACAAAAC AAAGCTGACA TTCCCGAGAT	11160
CATCATAAAC AGCACCACCG CCGGATAPAC GGCGAGTAAT AACGATTCCC TTCTCTCTAG	11220
CATAAGCTAA GTCAATTTCT TCATAAGCAT TTTGATTTCG GCCTAAAATA ACACAGGGTT	11280
CCTGAATATA AAACACACT AAGGGCTCTT CAAAGAGAG AGTGTTTAAT AGATAATCTT	11340
CCGTTGCTAA ATTTTCACGA AATCTCGTG ACTGCATTAT GACATAACGC ATAAGGTCGC	11400
CCCACTTCTT TTTCTCTAG TGTATCACTT TCTTTCAAAA TTGGTTCCTT AAACGTCTTC	11460
TCTGGCCATT CTCAAAGCG GATGATCAC AAAAAACGA ACGTGGGACA ACAATCACTT	11520
TGGATGTTTG TTCCACsCTC AATCTAAT AAACAGCAAT TGCAGAACTG TTTAGCTTTT	11580
TAATCTAACC AAGTACTAC TTCTGAACA TCTCGTCGGA CTGTTTCAA GCCGGGTGTT	11640
TGAGCTTTGC CAATCGGAAT AATCAGTCCG TTCACTAAAT GATCCTCAAT ATTAAATAAC	11700
TTGCGAACAG CTTCTGTTT APAACCAGAC ATCGCCACAG TATCATAGCC TCGATCTTTT	11760
GCAGAAAGCA TCAGCTGCAT CGCACTAG CCTTGGTCTA ACATTAGCCA CTCTCTTTT	11820
TGTTCTTCGC TAACATCATA GTACCAAGAA ACATTTTCAT TAATTCCTTG ACGAATTTCT	11880
TCCGTATCG TTCTTGTC GACAGCTTA GCACTGATT CATCTGCATT GGCCACTGTA	11940
TAAGCAGTCA AATCACCTAA TAAATATC ACCGCCGTG CCCCTTTAAC TGCTGGTTGT	12000
TGCATCGAAA CTTCGTAAAG GGCCTCTTT TGTCTTTTT CTGTGACAAT TAAAAACGC	12060
CATGAATTA AATTGGTACT ATTTGGCGG AACATCGCTT CCTCAATTAT GGCCATTAAT	12120
TCCTCTTTAG GAATCTCAAC ATCTGGCTA AAATCCCGAA CTGAGCGGCG GCTGCGAACA	12180
ACATCTGAT ACATCATATT TCTCTTCCT TCCTCTTGGC TTAATTGACA TAGTTATTAT	12240
TACTCACTAT GATGGAGAAA APAAGTACGC ACTTTTATT CACCTAGGAA GAAAAAGTA	12300
ACCATTACGT GTCACAAGGA GAACTATGA CAACTGATAA ACAAACAGT ATCAACTTAG	12360
CCCTTTCCAC CATTAACGGC AATGGAAAT TATCATTAAAT GGATGAATTA TTCCAAGGAA	12420
CCAAgCGCAA CGGTGAATTG ATGCGTGCGC TCGACGGAAT CACTCAGCGC GTCCTGACAG	12480
ATCGCTTGAG AGAATGGAA AAAGATGCC TAGTCCATCG AGAATCCTTT AACGAACTAC	12540
CCCCGCGTGT GGAATATACG CTCCTCCGG AAGGATATGC TCTTTACGAT GCCCTCTCTT	12600
CACTTTGCCA TTGGGGCGAA ACATTTGCAC AAAAAAGGC GAGACTTAAC AAATAAGTCT	12660
CGTCTTTTT TATCTATTTT GGGTATTATG GGTAACTTA AGATATTTTT GAAGTCCTTC	12720
TTGAATGAA TAAACGGGCG AAAAACCTAA GCCTTTCAA GCAGAACTAT CTGAAAGAGA	12780
ATATTTATA TCGCCATCTC GTTCCTTTTG ATAAGAACT GGCAAAGAAA GTCCCACTAA	12840
TTGGTTCAAT TCTTCAATTA ATGATTAG GGAAATGGCA ACTTCGGTAC CCACATTATA	12900
AACATGCCC AACGCTTCAC TTTTGGTGC TACTAAATTT AATGCAGTTA GGACATCCTC	12960

AATGTAGATA	AAGTCACGTG	TTTGCTGACC	ATCTCCAAAT	ATTGAAATT	GGCTCACTTG	13020
CCCTTGCTCT	AATTGGATAT	AACGATCCAT	CAAAATGGAC	AGGACACCAG	AATAGGGTGA	13080
AGCAGGATTT	TGATTGAGAC	CATAGACGTT	AAAGAAGCGA	ACTGCGCTCG	TGGGTACCCC	13140
GTGTAAATGG	TATTCATTGA	GAACATAACG	TTCAGCAGCA	AAC TTATCAA	TAGCATAGGG	13200
ACTCAAAGGG	CAAATCACCG	AAGTTTCTCT	TTTAGGCAAC	GTTGGCTCTG	CACCGTAAAC	13260
GGCTGCTGAC	GAAGCAAAAA	CAAAACGTTT	GAGCTTTTTT	TGATATTTTT	TTATTGATTC	13320
CAGTATTTTC	AATACACTTA	AAAAATTGAC	CTCATGGGTC	TCTAGAGGCT	GTGCCACAGA	13380
ACTTGCCACG	CTTGCCACGG	CCGCAAGATG	AAAGATATAC	TCAAACGAAT	GCTTTGAAAG	13440
TACTTCGTCC	AATAATTGCT	GATCCGTCAC	ACTTCCTTTA	ATAAAAACCA	ACTGCTCAGT	13500
TTGTTGCAAA	TTTGACACTC	GCCCCATGGA	CAAATCATCA	ATCACAAAAA	CTTGATACTG	13560
TTGACTATAA	AAATTAGCCA	AATTAGAACC	AATAAAGCCA	GCTCCTCCAG	TAATTAATAG	13620
TTCTTTCACC	TCTTTGCTTC	CTCTCTTTTT	TTTTAATTTA	ATTCCACTTT	TTTCGTGTTA	13680
TTTTCTCTTT	GGAATAAAAA	AATGCAATAT	CGAATGAACA	CATAACTTAC	AATAGCCACA	13740
GCTAATAAGA	GTTCAAAAGT	AGACAAAATA	TTTAACAGAC	TCGTTAGGCC	TACCTTTTCT	13800
GCAAAATAAG	GTTTAACTTT	GACTAATGCT	GTCGCAGAAA	TTACTAATGG	GAATGTAAAG	13860
GCAGCATAAC	TTGGATAGAA	TGTCACTTGC	ATAATTTTAT	ATAGATTAC	TACTACAATC	13920
CCATAAATAC	ATTGAGATAA	AATTAGCAAG	AAAACAACTA	ATGGCATATA	CAAATGACCA	13980
CCCATTGAAA	TATAGCCAGC	TAAACATAAC	GAACCTGGCG	CTGTCATAAT	CGTCAATAAA	14040
GGAATGGTCG	ATTCAGGCAT	TTCTTTTCGAC	TTCGTAATTC	TCTTGATTAA	AATTGGTAAC	14100
AAGAGTAAAT	AAAAGAACAA	CGCTACCCAA	AAAACAATTT	TTCCCAACTC	ATTGATAAAC	14160
AATTGAGAAG	TATTAGGAAT	AACGCCAATG	CCGACAAACG	TAATAAACCA	GCTTGGATAA	14220
ATATGTTCCA	GTTGATTTT	TGTTGGAAAA	ATATGCACCG	CTACAAAATA	AATCATTAAA	14280
ATAaAATGTA	AAATCAcTGC	GCCAATCCAC	AACACGTAA	TCCAAGCAGC	ATTGGTAAAT	14340
ATCCGATCTA	AGAAAACACA	GATAACCATT	AAAGCCATTG	TAAATGTTGG	TGATACCGAA	14400
GCAATAATTG	GATCATCTAA	GGTGCTTAAT	GCATGTTTCA	TTGTAAAAAT	AATTTTTTAA	14460
ATAACTAAAA	TCATTAAGAG	ACTGCCAAGT	ACACAATAAA	TGGTCCCAAT	TGTTGCATAT	14520
CCCAAGGAAT	AAAGTAAATT	CCCCAAAGAA	ACCATCCCTA	AAATTAGTCC	GCAAATGGGA	14580
ATCGGGACTT	TCCGCAAGAA	TTTTTTTCGCA	CTTTCTACCA	TTATGTAGTC	ACACTCCGTC	14640
TTTATTTATT	ATTATTGATA	TTAATCAAAA	GTATGTTATT	ATACTAAATG	AAATATTTAT	14700
ATTTGTAAAA	TAAATATTTG	CAATACCAAC	TATTAAGATA	GTAAATAGAA	AGAGGCGATT	14760
TTATGTTCGA	ATGGATCAAA	ACATTTATTG	CTGTGTATGA	AaCCAAAAAT	TTAGTACGG	14820
CGGCCAAACA	ATTGTATATT	TCTCAGCCAA	CTGTTTCTTT	GCAAATCAAG	AAATTAGAGC	14880
AACACTTTTC	GATCAAGCTG	TTCTATCGTA	ATGGAAAACA	ATCGGTTATT	CCTACCAAAG	14940

AAGCTGATTT	TTTATATCCA	AAAATGCTCT	CGATTATTGA	AAGCTTAACT	ACTTCTTTTG	15000
CGCAAGCTGC	AGAAAAAGAG	AATTTTAAAG	AAGACTGCAT	TATAGCCTCT	TCAAATACAG	15060
TTGCTATTTA	TTTACTGCCA	GATATTATGG	AGAGTTTGGT	TACCGCTTTT	CCACTCATCA	15120
ACTTTTCGAT	TCAACTGATG	AATTCAAACG	AAGTCGTGGA	TATGGTTCAA	AATAATCATG	15180
CCCACATTGG	ACTAATTGAA	AAACCCATTG	AAACCAAAAA	TCTACACAAA	GAAATTGTTT	15240
ATGAAGATCA	ACTAGTTCTA	GCAGGAGATG	CCGCTTCAAA	ATTTTGGCTC	TTAAGAGAAA	15300
AGAACTCTGG	CTTACGTTTT	TTCAATGAGC	TTTATCTCAA	CGAACACAGC	ATCAACTTAC	15360
CAATTATCGA	ATTAAACAAC	AACGAAGTCC	TGTTACAATT	GCTAAAAAAT	AACATTGGGC	15420
AATCAATTGT	TTCTCGATTA	TCTATTTCCG	ATGAGATCCC	TTGGCAACCA	ATTGATCAGT	15480
CCTTTGCTTC	TCGAAAAATT	TTCATTGTCC	AGACCGATCA	CCATGCCAAT	ACTTCCTTCT	15540
CTGAAGTTTA	TAACCGAATC	GTTGAAAAAA	TTCGTGAACG	ACAAACCTAA	AAAACAGGAG	15600
AACGACTTAG	CTTTCGCTAA	ATCGTTTTCC	TGTTTTTTAT	TCATCAGTTT	CCTCTACCAA	15660
TTCAAATTCA	AAAGGTTCTA	ACTCAAATAG	TTGGTGTAAG	ATTTGCTGGG	TTTCTTGGA	15720
CACTTGTTGT	GGCTCTTTAC	ACAAAAGAGT	GCTTTGAATG	GTCTGTTTTT	TTGATTGAAA	15780
AAGAATTCTC	GTTTCCTCTT	CCTTAAATAC	AACCTTGATG	GTTCCCTTTC	AAACACCGTA	15840
TATTCTTCGA	ACATTTCCAT	AACTAAATTT	TGACGAAACC	CTTCTTTGAG	TGCTAATAAA	15900
GTTACTAAAT	CTGCAACGGT	TAGCGkTTCT	AACGCAAnTG	kTTCaTCyTC	TACATGAACc	15960
AAGATAATnC	GtACACTTCA	CATtCGGACA	GGCATTGCAG	CCGCCATTAA	TCAAGTCTGA	16020
GGTAGTTGTT	AyTTTCACCT	GTTTCATCTT	AATCCTTCTT	TCCTACTTTT	CGAGAAAAAC	16080
ACAAGGGGTC	TCCTCAAAAA	TATTCAATAA	AATATGATTT	ACTTCTTGAA	AGACTTCTTC	16140
GTGATTCCGG	CATTGTTTTT	CCATTACTAA	ACGCAGATCA	TTCACAACAA	AAGTCAACGT	16200
CGTCGCTTGT	TGGGAAACAT	GTATACAACG	GTTGCCTTTA	ACAAAACATT	CTGTTTGTTT	16260
TTCAATGAAT	GGTTCTTTTT	GATAGCCATA	TGCTGAAGCG	ACACATTCCA	GGATTTCGTA	16320
GATTTCTATT	TTTGGAAAGT	CATAGCTACA	GCTACCGATA	ACTAGTTGAA	AACGGTTCGG	16380
CTGTTTTTCA	TTATTGCTTG	CATAATACAC	TGTTGTCGTT	GTCATTTTGC	CTCCTTATCT	16440
CCTCTATTTT	CCTACTCTTA	TCATAACAAA	AAAAGCACGC	TTCCCTTATT	TAAATTGTTA	16500
ATAAAGACTA	TTACAGATTG	GTATAACCAC	ACCTAAAAAG	TTCGTGATAT	AATTCACTAA	16560
GAAAGAAATA	TTACAGAAAG	AAGAGATGAA	CGATGAAAGT	ATTTAACAAA	AAAGTCGCAA	16620
TTATTGGTAC	TGGTTTTGTT	GGCACAAGTA	TTGCCTATTC	CATGATCAAC	CAAGGGATTG	16680
CGAATGAATT	AATCTTAGTT	GATATTGACA	AAGCCAAATC	TGAAGGCGAn	CAATTGACTT	16740
ATTAGATGGT	GtGkCyTgGG	GkCmAGAAAA	TGTAAACGTC	TGGGCTGGCG	ACTATCAAGA	16800
CTGCCAAGAT	GCCGATATCG	TCGTGATTAC	AGCTGGCGCT	AaTCAAAAAC	CTGGGCAAAG	16860
TCGTCTAGAT	TTGGTTTCAA	TCAATGCAGa	AATTATGAAA	ACAATTGTTA	ACAATATCAT	16920



GAAATCTGGT TTTGATGGAA TTTTAGTGAT TGCCTCAAAT CCTGTCTGATG TACTGACTTA	16980
TGTGGCCTGG CAAGCTTCTG GTTTACCTGT TTCAAGAGTA ATTGGAACTG GTACAACTTT	17040
GGACACAACCT CGTTTCCGCA AAGAACTGTC TCAACGTTTA GCGATTGATC CACGCAATGT	17100
TCACGGCTAT ATTATTGGCG AACACGGGGA TTCTGAAGTG GCTGTCTGGT CTCATACCAT	17160
GATTGGTACC AAACCTATTT TAGAAATTGT GGATACGACA GAGCGCTTAA CTAGTGACGA	17220
TTTACCAATC ATTTCTGATA AAGTGAAAAA TACAGCTTAT GAAATTATCG ATCGCAAACA	17280
AGCGACCTAT TATGGGATTG GTATGAGTAC TGCACGCATT GTTAAAGCCA TTTTAAATAA	17340
TGAACAAGCT ATTTTACCTG TCTCAGCTTA CTTAGATGGG CAATATGGGC AACAAGATGT	17400
ATTTACAGGG ATTCCTGCAG TCGTTGGCAA TCAAGGTGTG ACTGACATTA TTGAATTGAA	17460
TCTGAATGCC GCTGAAAAAG AACTCTTCCA AAAATCAGTG ACACAATTAA AACAAGTGAT	17520
GGCATCGTTA CAACCGAACG CTTAGTAATT TTTAACTAAA AATAACACTT CAACTAAATT	17580
ATTCTCTACC AAAATAGATT CTAATTCCCC TTTATCGTTC TTTTGGTAC CATGAAAGAA	17640
GAACTTCAAC TAAAGCAGAA ATTAGGAGGA CAAAAATGG TAATCCAAGG AGATACGTTA	17700
GAAATAGCG CACGTCGTTT ATTGCAAGAA CGTGGCGTAA CAATGAATGA TTTAGCTGAA	17760
CTTGTGTTAT TTTTACAAAA AGATTATATT GACAATTTAA CATTAGATAC TTGTTTAGAA	17820
AGTGTTGATG CCGTGTTAAC AAAGCGTGAA GTACACAACG CGATTATTAC AGGTGTCCAA	17880
CTGGATATTC TGGCAGAAGA AAACAAGCTA CTCTCACCGT TACAAGAAAT TTTAACGGAA	17940
GATGAAGGGC TTTATGGCAT TGATGAAATT ATGGCGCTCT CCATTGTAAA TGTTTATGGA	18000
TCTATCGGTT TTACGAACTA TGGGTATATC GACAAGGTAA AACCAGGCAT TTTAAAAGAG	18060
TTAAATGCGC ATGATGGCGA ACATGTCCAC ACATTCTTGG ATGATATCGT GGGCGCAATT	18120
GCGGCTGCGG CTGCCAGTCG GTTGGCCAC TCACAACCAG ATTTAAGTGA TATTACTAAA	18180
TAAAAAACTT GCCTCTTCAT TTTTATGAGA AGAGGCAAGT TTTTATTTG CTGGACTTCT	18240
CACAAGCTTT CTACTTAACA GTAATTATTT GCCGTGTTAT GTTACAGTCA AGAATAAACA	18300
GCTAAAGGAG GTTATCTAGA AATGAAAAAG TTTTACGCG TATTTTAAAT TTTTATTGGG	18360
GTATTCTTTC TTGCAAGTTG CGGTTCAAAA ATAGAGACAA AAACCTTTGT AGGTTCGCCA	18420
CAAGATGGGA TTGATTCCAC CTTAACCTAT ACGTACCAAG GAGATAAAGT CTTAACGCAA	18480
ACGGCCAAAA ACATTGTTTC TTATGACAAA TTGGGGATTA CAAAAGAAGA GGCAAAAACA	18540
GCGTTGGAAC CCGTTTCTAA ACAATATGAG GACATCAAAG GGCTTGATTA CAACTGACC	18600
TATGAGGACA AACAAGCCAT CGAAAAATTG ACGATTAATT ATGAAAACT TGATTATGAT	18660
AAAGCCAAAA AAGTGGATGG TATCCAGATT GATGGAGATT CTTCTAAGGG AATCAGCATG	18720
AAGAAAAGCC AAGAATTGGT GGAGTCACAA GGTTATACAG AACAAAAATA ATAAAAGAAA	18780
GCGTGCGACA ATGGGCATTG TCGCACGCTT TCTTTTTATT AATTACATTG ATGTTGCCAT	18840
CTTCACTGCT TGAATTGGGT CAAAATGATG GACAGAGTAA ACGACTGTAC CGGCTTGCCA	18900

AACTACTGAG	TTTCGATTAT	AGTCAGTGGT	TCCTGCCACG	TTCAAGTGTGA	TTTGACCAAC	18960
GGTATTCGGT	GCTGGTAATG	TTTCTTGTTT	CAAAATGTTG	ACAACATTTT	TCGCTAAATC	19020
ATCAGGCGAT	TCACCATTGA	TATTTGAGGC	CCGTACGACT	AAACTCCAAT	TGCCTTCTTG	19080
CCAAGATAAG	TAAC TAGAAC	CTGCCGCCCC	TTGTTTATAC	CCAGTAATAT	TGTAACCTAA	19140
ATCTACTGGT	TGACCGCCAT	TGTCAATGAT	TTGGTTCACT	GCATTTTTTG	CTTCGGCATC	19200
AGAGCCATAT	GTTTTTTTCT	CAAAAGAACT	AATTGGCGTT	AACTGATTGA	CACGTGCATC	19260
ATTCACTGGA	ATCGCTTCTT	TTTCTGCATA	ATATAAAACA	CGGAAATTGT	TTTGATCCGC	19320
TTGTGAAGTT	GTCGCTGCAG	TTAAAAAGTT	ATTCGTTTGA	CTCGTTAAAA	TTGCCTGCGG	19380
TAAGCCTTGA	CCAGGAAACA	TGTTTGCTAG	CTGGTTAAGC	ACTTG GTTG	CTTG GTCATC	19440
TGGATTCGCA	TTTTTCGCTG	GTGATTGCTT	CTCTGCTGAG	CCATTATTTG	CCGCTTCCTT	19500
CGTAGAAGAA	GCTGTTGCTT	TCGAATCAGC	TGTTGTTGTT	CCTGACGGTT	TCGCATCACT	19560
AGATTTTGTT	GTTGCTTTGC	TACTTGATTG	TTTTGTAGCA	GATGCTTTTG	TACTCTCGAC	19620
TGTGGTACTT	TCCGCAGCAC	TGGAACGACT	ATCCGTATTT	TCTGTACTTT	TTCCGCCGCC	19680
GCATGCTGCT	AAAGCAAGGG	TTAAAAGAAA	AGCACTAGAA	ATAAGCCCTA	TCTTTTTCAT	19740
AAGAATTCCT	CCTTGTTTTT	CCTACCCTTA	TTGTACCTTA	ATCATGTGTA	TCCGTCACCT	19800
GTTTGTGTTT	CGTTCTAAGT	TGGTTTAAAT	AACAGAAAGT	GGACGGTAGG	ATTCGAACCT	19860
ACACCTCAAC	ATTGAGCAAG	AATAAACTAC	TGTTAGTCTC	TATGAGTTTA	ACACCACGTT	19920
CATCCACCTA	TTTGATTGAT	CCTTTGTATC	ATAACTTAGA	ATAATTAAAT	TGCCTTTAAG	19980
GAGCTTCAAA	GATTATATGA	AGAATTGTAA	AAAATCTGTG	ACGAAAAAAC	GATTGAACCA	20040
ATCAGTAGAT	AACTGAGGT	TCAATCGTAA	CCAATTGTTT	AATTATCTGA	AAAAACTCAC	20100
AATTCAACGT	TAGTCCCTGG	TGCTTTTTTG	TTTCCCCTT	GTTGTTTTTT	TAAGAATATC	20160
CCCCAAATGA	CGGTGATAAT	GGGACAAAGC	AATGGGAAAA	AGGCATATGG	TAAGTATTCT	20220
AGTGCACCAA	CTTTTAACGT	TCCCATAATG	AAGGTACCAC	TAAC TCCCCA	AGGAATTAAC	20280
GAGTTGACTG	CCGCCCCAGC	ATCTGCCAAA	GTACGAGTTA	AATATTTTTT	ATCAATACCT	20340
AAACGAGTAA	ATGAGGATTT	AAATGTTTCC	CCTGGTAAAA	TAATTGACAA	ATATTGTTTCG	20400
CCAACTAGTA	AGTTAATTCC	CACAGAACTT	AAGGCAGTTA	AAGCGATTAA	TTTGCTGGA	20460
TTATTCACAT	AGCCTTTAAT	TTTATCAATT	AAAGTAGCCA	CGATGTTAAA	CTTAATCAAC	20520
AAGCCGCCCA	AACCTAATGC	CAAAATAATC	AACGCAGCAG	AACCTAACAT	GCTTTCAATC	20580
CCACCACGAG	ATAAAAGTGT	ATCAATACTT	TGATCCCCTG	TATCTGCCAC	ATAACCACTC	20640
ATTAAGATGG	TAGAACTTTT	CGCTAACTTT	AGGTGATGAT	CATTTATAAA	AGCTAAGATG	20700
ACCGCCACTG	TAGAGCCGAC	TAGTAACGTT	GGAATAGCCG	GTACCTTTTT	CCAGGCCAAT	20760
AAGAATAAAA	CCGCTACTGG	CAACAATGTA	ATCGGTGAAA	TCCAGAATCC	TTGATGTAAC	20820
GTTTGGACCA	TTGCATCAAC	CGATTGTAAA	TCCGCTGAAC	CAGATTGATG	CCCCAAGAAG	20880

ATATAGCCGA CAATCGAAAT AATAAACGCC GGAATAACAG TGTACATCAT GTTTAGTATA	20940
TGTTCAAAAA GGTTCACGCC ACCAATTCCA GCAGCTAAGT TTGTTGTATC AGATAACGGT	21000
GAAATGTTGT TCCCTAAAAA GGCACCTGAA ACAACTGCAC CAGCTGTCAT CGCATTGTCTG	21060
AATCCAAGAA TATGTCCAAT ACCTAGAAAG GCAATCCCCA TAGTTGAAAC AGTCGTAAAG	21120
GAGCTCCCTA CCGTCACACC AACTAGTGCA CAGACCACAA AAACGGTTGG TAAGAAGAAC	21180
TTCCTGAGA TAATTTTAAA ACCATAGACC ATAATTGTTG GAATCGTCCC AGATAAGATC	21240
CAAGTAGCAA CTAAGACACC TATCAGCAAG AAAATAATAA TCGGAATGAT CCCTGGTTTG	21300
ATCCCTTCCA CAATGCCCTC GTGGATCTCA TCCCAAGAAA ATCCTTTCCA TTTGCCATAA	21360
AGCAACAACA ACATGAATAC AAAAAGAATC GGTATGTGAG GGGTTAATTT TTGACCAATA	21420
ATTAAGTACC CGAGAATCGC TAACAAGGCA ATCAAATGC CTAAAcTTTC AGCGAACGAA	21480
ACAGTGGGTT TCGTTGATTT CATATGAAAA CCTCCTTTTA TGGAAAGCGA AACTGAAGC	21540
AATAATCGAA CTTGCTTCAG TTnCCAAACG ACAATTTACT ATAGTCATCA GGTGAAAAGG	21600
CACTCAGCAA AATAATTGCT AGCCTTTTTA TTCACCTAAA TCTTCTTTTA CTTCCGTTGC	21660
CCAACCTGCT TTTTAAAGC GATTAmTGAT TAACGGAATC ACAAACATAA CAATAGTGAT	21720
TAATGCAACT AAGACTGCAT CAGTCACGGT TGTGTACCT AATGTGGCGA AAATAACAAT	21780
GATAACACTA AAAATTAACA TTGCTGCCAC TAAATAAGCT AAGCCATTAC CTTTTTTACC	21840
AATTCGGTAA GGACGTAGTG TATCTGGTTG AGATTACGT AATTTAATTA GGGAAATCGC	21900
AATCAATACA TACACAATAC TATAAATAAT TGTTGTTGCA TTTGTTAAAG TTAAGAATAC	21960
AGCGTTGACA TCATCCATTA CCCCATAAAG TAAAGCGAAT AATGAGATAA CCACAGATTG	22020
TGtTAACACA ACGTTACGAG AAACGCCATA TTTATTTTCG CGATGGAAAC CAAATTTTGG	22080
TGGCAATAAG CCATCACGAG CAACTTGAAT AATTGTTTTT GATGGTCCAG TAACCCAAGC	22140
AGATAATTGT AATAACACAC CTAAGAAGAC CATGAACTA AAGATATTAC CAATAATTGT	22200
TGGTAAACCT AAAACGTCAC AATAAAGCAA AATTGGTTGT GTAATGTTAG ACAATTCCAT	22260
TTTGCCATTT GGAACAATAT TAGCAACAAA CATTGCATTA ATAACGTTTA ATAAACTAA	22320
ACCAATTAAT GAAATGAAAA CGCCTTTTGT ATAATTTTTT GTTGCATCTT TCAAACGAGG	22380
AATATAAACA GAAGACATTT CAATCCCAAC AAAGATAAAT GCAATCCCTG CTAGATATTT	22440
AAATGTGTCC ATATTTTCTA AGTTTGGTAA AACTTTTGAC CAAGAGAAGG CACCTAAATA	22500
GCCGCCAGGT GTTAAGCCTA CTTTAATCAT TGAAAGAAGA CCTAAAACAA ACATAACTAA	22560
TACTGGGATG TAAACACCTA GCCAAACACC AATATTTCCG CCAACTTTTCG CCATGTCAAA	22620
TTTCAAGTTC AAGATTGTAA TAATCCAGTA AATCACTAAG ATACAACCTA ATACGAACCA	22680
ATGATTTTCA CCTAAATGAA CATTaCCAAA TGTATTaCCT AGCAAAGGGC CTAATGTAGA	22740
AGCAACCATT ACCATTCTTG GGAACATTTG CACCCATAGT AACCAAGCGA CAACAAAACC	22800
CCATTTGCTA CCTAAGGCAG TTTTAACCCA AAGTTGAGGG CCACCTTCTT GAGGAAGCAT	22860

CGTAGATAAT TCACCAGATA TCAATGCGAT TGGACCAGCA AAGAA TAAGA CCGCGAAAAT	22920
TGTGTAAGAA ATTAATGTCC AACCAACAGC AGCAAGTGTT GGGATACTAC GTACAGTTGC	22980
ACACAATGCC ATTGTTAAAC CGATAAAAGT ACCTAATTCA ATTTTTTTAA CTTGTTGTTT	23040
ATCCAATTTT GTACACTCCT TATAAGATTT TGGATAAGGA GCATGAATGA CTGTTATACG	23100
CTGAGTAAAA CAGTCAATCG ACTGCCTCTT ACCCAGACAA AGAATTGGTT AAGAGGTTAG	23160
CCATCTGATT GATTATCTC TTAAATAAAT CAATCAGACG AACGTTCCCT CATTCCGTTA	23220
CCATTTCTTT TTTCGGTTTA AGATTCAATT ATTTTACGTC GTAAATTTGT TCTAATTTTT	23280
CTTGTAAGC AGCTTGAATC TTAGGAGCGT AAACATCAAA TTCTTCTTTA TCGTTCATGT	23340
ATGGAGTCAT TACAGCAGCA CGTAACACAG TTACTTTACC TGCACGGTTC CATTCTTCGT	23400
CACTAAATCC CAAGCTGTTT ACAAATTTTA ATGGGCTGTT GCCATAATCT GGAATAGCAA	23460
AGTCAGTATG TGAAGTAATG AATTCGTTGT TGTAATATT TCCTTTAACG TATGATGCAT	23520
AGTCATATAC ATCATGGTTT AATTTGTTCA TTGCTACTAA ATCATCGTTG CCTTTTTCTT	23580
TAAACACATA GTCAACCATG TTGAAGCTG GATGAGTTAA TGTGTGGACT TCGATTTCTT	23640
TGTCGCCAAC TTTGAATGTT AAATCATTCA AGAAGTTATA GAAATGATGT GATCCTTCAA	23700
TTGAAGCACC AATTAATTTA CCATAACCTG CAACGTTTAA AGGTAAAACG TGATGTGCAG	23760
CCCATACAGA AGCAGCTGTT GCACCAGCTT TTGAACCTC AAGGATATAA GCACCTAGTA	23820
ACGCAGGAAT ATCGGCACCT TTTTCAAATA CGTAAGTTGC GAAGTAAGAA ATAACGTCGC	23880
GCATACGGAT ATCTTGGATA ACGATACCAC CAGCTGAATA AGGGATATAA CCCATTTTAT	23940
GAGGGTCAAT TGTTACTGAT TCTGCTAATT CGATTGCTTT ATATGCATCA TACACTTCTC	24000
TTGAAATGTG TTCTTTTTTC TCTTTGAAGA CACCGTATTC TTCGTGAACA TCTTGTAAT	24060
CTTCGTAAGG GATGAAGTTG TTGTCTTCGT CTAAGAAGAT GGCACGTCCA TAACCACCAT	24120
AAGCAGCATC AACGTGTACA TAATAGTAAA TACCGTCTTT CATTAATTCA TCGCGTAAAG	24180
CAATAATTTT ATCGATTGAG TCAACGGCAC CTTCTTCAGT AGAACCAACT ACACCGACAA	24240
CACCTAGCAC TGAATTTGT TCTTCTGCTA AGCCACGAAC GATTTTTTCT AATTCGTGA	24300
TATCCATACG GTAGTTATGG TCAACTGGTA CTGGGATAAC TTGATCTAAA CCGATACCGA	24360
TAATATCAGC AGCTTTTAAC CAAGAATAGT GTTAGTTTG TGGAATAAC CATTTACCGA	24420
TTGCTTGTA ATGTTTACCT GAACGAGCTG AATGAGCTTT GATTCATCA ATTCATCTT	24480
CAGCTGATTC TAATAAGTCC ATAATTTCTT TTGTTGGCAT GTTCAATAGT TCCCAATCTG	24540
ATTTGCCAGC AACTAATTCT GGTTTTACTT CTTTCATAGC AAATGGTAAT GATTTAATGT	24600
TACGGGCATA CCATAAGCCT TCTAAGTTAG CTAAAGAACC ATCAGCAACG ATGTGTCCCC	24660
AACCATTTTT GTAGCTCATT AAGTGAGCAA ATTCATGTCC TACTTCTTCT TCCATTTGAG	24720
AAGTTGCTGG AGAAGATTCA TAGGCAACGT TGTTCCCGTT CCATAGCATT GCAAAGTTGT	24780
AAGCTAATAG AGAAGGCATT AATGTTTCTG AGTTCATATG TCCCAATAT CTACCTGCTG	24840

TATGCCATGG	AACTGAATGT	GTACGCATGC	GAGAAGAGAT	TTCATTCAAT	ACATCTTTCA	24900
TATGGTTGAC	AGTTTTTTTCG	TAGCTTTCAG	ATGTGCGTTC	TTGAGAAGAG	ATAACTGGCA	24960
TGTCTTGTGG	CATGTAGTTT	TGACGCCATC	CTAAATGTTT	ATCTACTAAG	TCGATCAACA	25020
AGTCTTTATA	TAATTGTCCG	TTTTCGGCTT	TGTCCCCAAT	GAATAGTGCA	TTAAGGTTCA	25080
TTTCGCCTTT	TGCTAATTTT	TCGTTTktCA	TACGAAAATT	CCTCCTTACA	AATTTTAATG	25140
AATTAGATTT	TGCATTACACA	ACTTACACCC	AAACGGCTAA	AGTGTAAGA	AGAAaAATAA	25200
CAATGGtTAT	CATAGCTCTT	TAATTTTATA	ATTTTTTTCAC	AAGGTGTCTT	GTCaGaACCA	25260
TTATATaAAA	AAATAAGAGT	GGGaTTgTGT	GtACAATCCC	ATCTTATTAG	GTGTATTAAC	25320
GGTTTCACCA	CTACCTGTAA	GAGnAAATAG	AATTTTAACT	TTGATCAAAT	AATCACACGT	25380
CTTCTCTTCG	TGAAATTATT	ATCAATTTCT	AATCCATTAC	TTTTGCTAAA	AAGTAATTTT	25440
TCTtCCmTTT	CCGTACAACA	ACGAATTTtC	CATCGAAATG	TTGTGTTGGA	TCGACTGCAA	25500
AATCTAAATC	GGTAACTCGG	TCACCGTTAA	TACTTATAGC	TCCTCCGCTC	ACATCTTCGC	25560
GAGCTTGACG	TTTTGATGGT	TCGATTTCTG	TGTCAACTAA	CCAATCTACG	ATGTTTTTAG	25620
CATCTTTGGT	AATTTCAACA	GTTGGCACAT	GTTCTAACCC	TTGCTCGATT	TCCTCAATCG	25680
TTAAGTCTTT	AATGTTGCCT	GAGAAAAGAG	CTTCTGAAAT	TTTTTGCGCT	TCTTCTAATG	25740
CTGCATCATC	GTGAACAAAT	CGTGTCACCT	CTTCAGCTAA	GCGTCTTTGT	GCTTCACGTT	25800
TACCTGGTTC	TTTTTCAACT	TTTTcAGTAA	CGCATCGATT	TCTTCTTTAT	CTAAGAAAGT	25860
AAAGAATTTT	AAGTATTTGA	TGACATCACG	ATCATCTTGG	TTTAACCAGA	ATTGGTAGAA	25920
TTCAAATGGT	GAGGTTTTCT	TAGGATCTAA	CCAGATAGCG	CCACCCGCTG	TTTTCCCAAA	25980
TTTTGTACCA	TCTGCTTTTA	GCATTAAAGG	AATGGTTAAC	CCGAATACTT	TGGCTTCTGG	26040
TCCTTCTTTT	TTACGAATCA	AATCTAATCC	TGCAGTGATA	TTGCCCCATT	GATCAGCGCC	26100
ACCAATTTGC	AATTGAATAT	TATGTTTTTT	ATGCAATGTG	TAAAAGTCAA	TTGATTGAAG	26160
AATTTGGTAT	GTGAATTCTG	TAAAGGAAAT	GCCACTTTCT	AAACGACTAG	CGACAATGTC	26220
TTTTGCCAAC	ATCGTGTTGA	CATTAAAGTT	TTCCCGTAA	TCTCTTAAAA	AATCTAATAA	26280
AGATAGTTCT	GATAACCAAT	CGTAGTTGTT	CACCATTGTT	ACCTCAGCAT	CTTTACCAAA	26340
TAATTTTTTTC	ATTTGGTTTG	AAAGACTGTC	CACATTATGT	TGCACAGCTT	CCATCGTTTG	26400
TAAGACACGT	TCGGTTGTTC	GTCCACTTGG	GTCACCAATT	GTTCCAGTTC	CGCCACCAAT	26460
TAAAATGTAT	GGGTGATGAC	CTGCTAATTG	GAATCGTTTC	ATCATCATAA	AAGGAATTAA	26520
ATGTCCAATA	TGCATGCTAT	CTCCAGTTGG	ATCGACACCG	CAATATAGCG	AAATGCTCGT	26580
ATTTTCTGTA	AGTTCTCTTA	GTCCTTCTTC	GTTTGTTTGT	TGATTGATTG	CATCACGCCA	26640
TGCTAGCTCG	TCAATGATAT	TCATGTTATC	ACCCTCCATA	TTTATAATTA	TCTCTTTATA	26700
GGAAATCGTT	CATTTAACGA	AACAAACGAT	TAACACACTT	GTTTTTACAC	GATATCCCGT	26760
GTTTATTTCT	ATGTATACAC	GGGTTACGA	AATTTTCTTT	GGTCTGCCAC	GTATTCAAAA	26820

AAAGAATACA AAAAATCCCT AAGTCGTCGA AACGACTTAG GGACGAATTT CTCGTGGTAC	26880
CACCCAAATT TATACATAGT TTAGTTTATT TAAAATACTA TATACCTCTT AACAAATAAC	26940
GCTGTTACCC GTGCCCTGTA GGCAGAAGCT CAAGAATGTA ATTCGATAAA TAACGTGTAC	27000
TAGTTTCCAG CGCCACTAGC TTTCTGTAAC AGTGGTTATT TCTCTACTGC TTTTCTGTCA	27060
TTGCTTTGTA TTAACTTGT ACATGTAGCA TATTCCTTTT CGTTTTTTTT gTCaACACTC	27120
TTkGCTTAA tTTTktTACT TTTATAAAAG ATGTCCCTTT TTCACAAAA TAAAATAACG	27180
AATTCAAAAT AATATTTATT TGTTTTAAAA GCAACTACAG ATAAAAAAT AACGAAAAA	27240
TTTAAAAAT ATTCTTTTAA ATTTTTTTGA CTTTTTTTGk GATTTTCTAT ACTTTTATAA	27300
AATTCCCGAr GAAAAATGAC GAACTTCGGA TCAGAAGTTC GTCaTTTTGT GAAAaaCAGT	27360
TTTAAAGAAT TGGTGAAAAT AATCGTGAAA TTTGTGTGTT AAAAATCAAC CATTTTGACA	27420
TGTCGCGAAC AACTTCTGGA GTCATTGTTG TACATTTTTT TCGATCTTCT AAGAAAATTG	27480
CTTCTAATTG ATCAATGATT TTTTATCAT ATAACACAGC ATTTGCTTCA AAGTTTAATT	27540
TGTAACTTCG GATATCTTGA TTGGCTGAAC CAACTGTGCA GACTTCATCA TCCATGATCA	27600
TTGTTTTCGC ATGTAAGAAA CCACCGTTAT AAATTAAAA TCAATATTT TCTTTCATCA	27660
ATAAGCGGCC GTAATACTGT GTTGCTCGAT AAATAAACGG ATGGTCGGGC TTATCTGGAA	27720
TCATAATTTT AACATCTACT CCTGAAGCCG CAGCAACCTT TAAAGCAGCC AAGACACTAT	27780
CATCAGGAAC TAAATACGGC GTTTGAATCC AGACTCTTTT CTTGGCAGAA GTAATCAATT	27840
TAATAAAGGC CAATTAAATT TGTTACGGT CACTGTTAGG ACCACTGGCA ATCATCTGAA	27900
TGGATGTATT TGCTCTGGC ACCAATGCTT CAAGTTTAAA GAAATAATCC AATTGATACG	27960
CCACACGATT TTTCTCGGGG GAAGAGACGT TCCAATCCAT TAAAAACGC ATTTGCAGTA	28020
ATGAGGCTGC TGGCCCTTGA ATCCGTAAAT GTGTATCCCG CCAATAGCCA AACTTTTTAG	28080
TCGTTCCGGC ATATTGATCG GCAACATTAA AGCCGCCAAT GTAGCCTACT TTTCCGTCAA	28140
TAACAACAAT TTTCCGGTGA TCATGATAAT TCAAACGAAA CTTCAAAAGT GCTTTTTGAG	28200
AAGTAATAAA CGTTTGGACA AAACCACCGT TTTTTTTCAA TTCATTCAA TGATGAACCT	28260
TGGTGCCCTT TGAGCCAAAG GCATCGTAAA GCAACCGCAC TTCAACCCCT TCAGCCGCCT	28320
TTTCTTCTAA TAAGTTCAA ATTTTGGTGC CGATATGATC CGTTACAAAG GCATAATATT	28380
CAATATGAAT GGAGTGCTGA GCTTTTTTTA TATCCGCCAT CAACGCATTG AATTTTCCG	28440
TACCATCCGT AAAAATTTCA ACATCATTTT TCTTAGTCAG AGGCATCCTA TTAAACGAGG	28500
AAAAAAAGTC TGTAACCTGT TGTTGCTCCT CCGTCGGCAT CTTTGGACTA TAAAGTTCAA	28560
AACTTTCTTC TTGGAATTC TGGAAATTCT CTAATCCCG TAAATCACTT TGTGCAAAT	28620
AAAATTTCTT TTKGTCGGkT AAACCTCGAC CAAAGrATAA ATATAAGAAA AAACCAAAAA	28680
CGGGTAGTGG	28690

## (i) SEQUENCE CHARACTERISTICS:.

- (A) LENGTH: 6146 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

AAATTATTGA AAAATCACCG AAGCAAATC ATACAAAATA TACAAAATGG CAAACAGAGT	60
CCATTATATA ACAAAAATA AAAAGTCGTT TAGACTTTTT ATTAGAACAA AGTTCTTCAA	120
TTAATGACTT CCTAGAAAAA GCAGAAGCTT TGAATTTATC GGCAGACTTT TCTAATAAGT	180
GGACGACCTA TCGATTGCTT GACGAACCAC AAATAAAAAA TACTCGTAGT CGAAGCTTGT	240
CAAAAAGTGA CCCCCTAGA TATAnTTATG AAAAGATTAT TGAACGATTG AAAGAAAATA	300
AAAACGTTCT AACAGTGGAA GAATGTGTCA AAAGtACGAg AAAAAATGA-ACAAGAAAAA	360
AATAATTTTG ACTATCAaTT TACmATAGAA CCaTGGCAAC TGTCTCACAA AACAGAAAGA	420
GGATATTATA TAAATATCGA TTATGGCTAT GGCAATTCTG GAAAGCTCTT TGTAGGAGGG	480
TATAAGGTTG ATCCATTAGC CAACGGTAAT TACAATATCT ATATTAAAAG AAATGAGTAT	540
TTTTATTTTA TGAGCGACAA GGATGCTACT AAAAGTAAAT ACTTAACAGG GGCTTCGTTG	600
ATTAAACAAT TACGACTGTA TAATGGACAA ACACCACTGA AAAGAGAACC TGTGATGCAA	660
ACCATTGATG AATTAGTTAG TGCCATTAACT TTTTtagCGG CGAATGAAAT CGAGGACACA	720
AGACAACATA AATTATTAGA AGAAAAATTA GAAGTTGCTT TTCTTGAAGC AGAAAAAACT	780
TTAGAGACTC TTGACGAGAA GATGTTAGAG CTGCACCAAT TAAGTAACCT CTTATTAGAA	840
AATGAAGTGC AGGGAGATCC AGAGTTAATT CAGAAAAAAT TGAAAACATT GTTACCGGAC	900
GCAACGCTTG CGGAATTTTC TTATGAAGAT GTTCGAGGAG AAATAGAAGC AATTAAAACC	960
AGTCAAACCT TGTTAGAGAG TAAACTGGAA CGAACTCGAA ATGAAATTAA TCAGCTTCAT	1020
GAAATTCAAG CCGTACAGGA AAAAGAGCCA GAGAAGCAAA CCGACATTAA GCCCAAACTA	1080
TAAACTATCT ATTGGTGAAC GGAGAGGAGA ATTTTATGTT AGTTTTTAAA CGTATTTTGT	1140
CTGTAATACT ATTTTTACCT TTTTTTATC TTCATTTACT TTTCCGATAT ATTGGAATTT	1200
TAACATTGCT TACTAGTGGA TTTACCACCA CTTTATTATT GATTATTGCC CGAGTTTGG	1260
CTTtCTTTT TTTAATAACA CCAATTGTGA TCTTAGCAAC AGAAGAACTT AGAATCGAAT	1320
GGGCAAGTTG GGTGGTTGG CTTTATTTG GATTACTTGT TCTGCTTTT CATTATCTAG	1380
CTGCCTTGAA TGAAGAATTA GGAGGCTGGT TTTTAGATTT TTCTGATTGG TTAAGCAAGG	1440
TTCCTAGCAA ACTGTGGGCT GAATAATTGC TCCTATGCTA AAACGCCTTA ACTTAAAAA	1500
GCCAAGGCGT TTTATGGTAT AATATACGAG ATAGTAAAGG TTTATGACGG TGGCTAATCT	1560
CTGAAAAAGG GGTGATTGCC TATGTTGATT TTTGGCTATA CTCCTAGAAA GGACTAGCTT	1620
TGACTATTTT TGAAGCGTTA ATGTTTGC GA TTTTCTTGC GACATTAATT CTAAAAATA	1680
GTGATAAAAA CAACAaAAAA TAACCGTCAA CTTTAGCCGA GTTTGGTTAT TTTTtGTAA	1740

AATATTTAAC TATCGGTCAC CGTCTTAAAC GGTGCTATCA TGAGGGCGTA TGACTAGTAC	1800
GTCCTCTTTA CATGTTTATT ATAACACATT CTCTACAAGA AGAAAGGCGG AAGCCTTTCT	1860
TTTTTTTCAG CGCGCCTTAA CGGCGGTGGC GGTGGGAGTT TGGcGCGCTG AaAAAAaACGC	1920
TCtTTTCTTG ATTTCTAATC AAGAAGATAC TCGCTAAAAG CGAGTATCAG GTTGTCTTA	1980
TATTATGATG ATAAGGCGTG CTGTAAATTC TCTTGCCATT GTCTGACTGC TTGTTTTATC	2040
TTTTCTGGTA CTTsCTGACG GTTACATwtn CCTTGATTTT GTATAACCTG TTGATTTCTGA	2100
TATTCTAATG TGAAGTAAGG CATATCTGGC TTTTCTTTTC TTCGGATAAA AACAAATCGTT	2160
GTTTTTCCTT GCGTATGCCG TTCGATATAG TGTTGACTGC CCACACAATG GCGTAGAATA	2220
CTTCCTTCTT GAATGATTTT TTGTAATGAA TGTGGTGTGA GAAAAAGCAA ATCGTCAATT	2280
TCAGCTTCTA ACGCTTTAAT TTGGTTTTTC CGTTCTTGAT ACTGCTTTTC TTCAATCTCA	2340
CTTTTCAAAA GGTTTAGCGT GTTCATAGCG TTATCGTGAG CTAATTGTAA GTTTTCAGGA	2400
TACAATACAC TATCGTTATT AAGTGGGGTG TTCAACTCTT CTAACATGTG TAAATAGTCC	2460
ATATAATAGT CAAAACGCTC CCCTTTTCGT ATAAACCATT TTTGAAACTT TGTTAAGTTC	2520
ACTTCTTTTCG GTAGTTGCTT AATTTGCGAA TAATGCACGT ACTTTTCTAT TTCGTTTATC	2580
ATAGGCGCAT TTCGCTCTTT TAACGTTTTT TCTAACATGA ATTCATGAAA GGTGGGATTT	2640
CTTGTTTTTA AtTTTTGtTT ATtCGCTTTT AGCCATTTCT TAGTAATTAC TCTCATATCT	2700
ACTTTATGAA AATACAGTCC AAAAACGTTG CGATAGTCGT CAAAAATGAT TTCATTGCGT	2760
AAGGTTGTGG CACCTATCTT TTGTAAAAAC TCAATTTCTG TTCGGTATTT GTAAATATGA	2820
GCTAGTTCTT GACGTTCTAT TTCAGGTAAC TGTAAGTACT TTAATCCGA TACAGTGGTT	2880
AGTTTTTCTT GCCAGTTATT TGGATAGAAT ACTGTATTTG TATAAGCACT ACTCATAAAA	2940
CCATAATTTG CTAACAAGCC CTCATGCCAA TGTCGGTTCT CATAATGGGC TTTAACATGG	3000
GTATCGTTTG AAAACCGTTC AAGGTTTGAC GGCTCATAGA TAAGGTTTTT TATTCCATCT	3060
GTAATCGTCT GCCAATAACT ATGGGTTTGA ATTTCAATTC GTTTTGACAG AACGAGAATG	3120
ATTCCAAAGG AGTAAATTT CGTTGGAAAC GATAATCGTG AGTATTTTGT TAATCGTTTT	3180
TTTATGATAG GGcAATTCTT GCGAGAACTC GCAAGAATTG TCTCCTTTTT GTTTTGCCAT	3240
TCATACGTGG GTATTTGTGC AGTACACCAG TTAaaaaAGG CTTTGGGTGG GGTTAATGGT	3300
TTTGGTGTTT TTTTATTCAT ACGCACTCGA ACAAATCAAT TTGTTGCTCC ACCTCTTTTC	3360
TTTCTGATTT TGTTTCTGAT TTTTGGTTG GTTCTTTTTT TGTTGTGGTC GTTACTTTAC	3420
CAACTGCTTT GGTTTCTGGT AATTCTCTA ATAAGAAATA CTCTTTTATC CAACTGAAAA	3480
CCACTTCGTC CGCCACCATG GCGCTTGATT GGTTTTTGGC TTGTTCTTGG GCTTTTCCTA	3540
TGCAGTAAGT CATAGCACCT TTTAACGTTT TATCTTyTTT GAGAATACCT AGAAAAAGCG	3600
CTTCGTCTGT TTGTCGGCAT AACCAGTTAT GAACAACCTC TAGTGACTG TCATGGTCGG	3660
TTTCTAGTTC TTTTArcATT TTCTCAAAAG CTTGTTCTTG CGTGCTTTCT TGAGCTGTAA	3720



AGAAAAATCG TTCGTAAGAA wTTTTGTGCGC TAACTACTAG CCAyTkrTAK TTrTTTCCTT	3780
GGTCTATyGT TTCGTCAATy GTTAGAAAmA STGAYGArTA GTkwCCGTtG GggTCTTTwA	3840
TCGyAACrAT yGCTTGTTcW GGTAAATTTT CHAGGTAAgy CArTAACGTT TCGACTGGAT	3900
AAAGAGCGAC CTCATAATTA GSATAAAAC ATAAAAACTG AACGTTTTTT CCGTCTTTTT	3960
CAATGTGTTc TTTTCTGAAA GCGACACTGA TAAAGCCTTG TATGTCGTCA AAAATTGTcG	4020
TTGCTAGTGT TGGTTTGTAT TGTGTGAGTG CCGTGTAA TTCATTTACA GTTTGTATAT	4080
TTTGAAAGTC CATAACGTTc CCTCGTTTCT TTTAAGCTTC CCCTCTTACA TATGTAAGAG	4140
GGGAGCAAAT TTTcGCTTGT CAAAGGTcAG TCGGTACGTT ATACTTGAAC TACGGTTAGT	4200
AATTCAAGTA AAGTCCTATT AGAAATCTTT TTTcGTTAGC AAAAGGTTTc TTTTTTTGTC	4260
TTTTTTTCGT GGAAATACTT TGATAATGGT TCAAAATCTT TCATATCTCC ATTGCAATTc	4320
ATAGTACTTA CTAGCCACTc TCCAGACTGA CGTTGTGCAT ATAAATAGCC GTTGTTATCA	4380
GAGCCACGAT AAGGTGTAAA GGTTACTTTC CCTTCAACGA GTTCCGTCCA ACCGTCAATT	4440
CGGTCAAGGT CAGTTAGCGT GAGGTATTGT TGTCTCTTA TTTCTTCTAA TGATTGTGCA	4500
AGATACATTT CTGTTTCTGG TCGACATAT AGACAAAACCT CGTGCCAACC ACAACAAGTT	4560
AGTTCCCGAC AGCCGACATT GTCGTATTcG TAAGTACGT TTTCTTTCAC ACCTAAACAG	4620
GACAGGCCTT TTTTCTGATT GATPACTTTT TCAATTTTTT CAGGGTCTTG GTAATGTTCA	4680
TATAAAACAG CACCTACTCC CTCTATATAA CCGTCCCAGT GACAATATAT GCCTTGATAC	4740
GTTCGGTCTG CTTGTTCTTT AAAPATTGCT GTTCGTGTTc CCATAAAAGC TCTCCTCTTT	4800
CGTTGTTTTA AATGATAAAC TCTGTTGCTT CGTGATTATC TTCTGTCAAG ATAATTCTT	4860
GGTCGTAATA CATGTTAGAA ACCAATTCAT GTGCCTGTTc TTCGCTTTCA GCAAATACAG	4920
AAACAGCTAC CTCTAAGGTT TCAACAATTG TTACACAGTA TTTTTCTTT TTTTCTGCCA	4980
TATTTATACG TCCTTCCATT CAATGTGATT TAAPAGAAT GCTAACTGAT TCAAACGATT	5040
ATTGGcATTT TTCAGAcAAC CATAGTTTAG TAACAATGTA GGCGTATCAT TTTCAAGAAT	5100
GGCTTTATTT AACTTCTGAT ATTCTTTACG TGCTTGCTCT AGCTGATTTT TTAGCTGTGG	5160
GTCGTTcATT TTTACTACTT CCTTCTGGT ATTCTATTAT TTTGGTAAT CTCTATCAGT	5220
GTTTTTTGCG CCTCTAACGC AGTTAAATTT GGTAAGCAAT CTCCTAAAGT CATACCGTTA	5280
CCGACAATTA CAAAAGGTTG AAGTAACTGC CCTACCGTCA TTTTTCGTC TGAATGAAAA	5340
AGAGTAAATT CTTTTTcAA AATATTTTCG CTTGTAAACA AACATTCAAT ATACACTTTG	5400
CTAAAGTCGT AAAAActTTC TATTGAaAT AATTTTGTAA GATAATATTC TACAGTTTCA	5460
AAATCAATGC AATCAAAATA CAAATTATTT TCAATCAATCA AAAACATTG AGAGCGTAAT	5520
GTGTTAAATA AATCTAAACA TTCTTCTAAA TCTAGCGATT TTAAATCGTT ATTTAAATAA	5580
CGATTAATGG ATAAGCTATT AATGAAATTT TCTTTTTTGT TCATTGGTTC TCTCCTTTAT	5640
TTTTTTGTTT TTGAGCAACC ATTTTTGTCT GCGTGAGGG AGGGAAGCTG TTCGCTTCTT	5700

TGGCTCATGC TAGGACGTGG GACACTATCC TCCTTTCGTT TTTAGATTTG ATTGTTTCATT	5760
TTGCCAGTA TGTTTCGGTC GGTGCTGTG CCGAGAAAGA ATCATAGACA CCAAGAAAAG	5820
ACAAGGGCGA ACGGAGTGAG TTCATATACC CTTGTCTTTT CGACTGGCTA TGATACTGTA	5880
ACAAACAGCG ACTGACCGGG ACATACGACA ATGAGCGGTC ATAAAACGAT TGAGAGGACA	5940
CACGTCCGTA AAAgCATGAG CCAAAGAGGC GAAATTTTAA TGCACTTAAA AGAACTAGAC	6000
GCAAAGCGGC TAGTTCCTTC TTTTGTTTTG CCTTGCTTTA GCAgGCATTT CTTTCGAGCG	6060
GCGGTGGCGG TTGGAGCGwG ATAGwAATAA ACACTAAAcA aAGTAATnAG GTATGAAAAA	6120
TGGAAAGACG ATTTTTCAGT GAATAG	6146

## (2) INFORMATION FOR SEQ ID NO: 140:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 656 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

TTTTTGTCAT GATTGTCAA TACTATGCTT CAGTCGCTTC GCAAGGTTTT GGGACGGAGC	60
TAAGAAATCA ATTAATGAAG AAAATTAATC AGCTTTCACA CAAAGAATTG AATAGTTTTG	120
GTACAGATAC CCTCATCACC CGGATCACAA ACGATATCAA CCAGCTTCAA TTAGCTTTAG	180
CGATGTTTAT TCGGTTGGTC ATTCGGACAC CTTTTTTAAG TATCGGTTCT GTGGTGATGG	240
CTTTTTACAT TGACGTGCAG ATGGGCTTTC TTTTCTATT ACTTTTACCA ATTTTtagcc	300
TTATTCTCTT TATTATCATT AAAGTGACTG TGCCTTTATA TCAAAAAGTC CAAGAATATT	360
TGGATCGGTT AAACCGTCAA ATCAGTCAAA ACTTAAGCGG TGTCCGTGTG ATCCGTGCGT	420
TTGCTAGAAA GGAAACAGAG CAACGACATG TTGATAAAGC TTCAGATGAT TTAGGCGATA	480
TTTACATTCG TGTATCGAAT GTsTCTGCTT TATTAACGCC TTTAACCACT TTGATTATGA	540
ATGTTGGAAT TTTATTTTAA CTTTATTTTA GTGGCTTAAA AGTTTCTTTT GGTTCCTTAC	600
AACAAGGGGA AGTTTtagCA TTGATCAATT ATATGAATCA AATGATGCTC GCTTAA	656

## (2) INFORMATION FOR SEQ ID NO: 141:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2776 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

CGGAAGATTG ACAAAGATC CTGACTTACG TTACACCGCT AGTGGTTCAG CTGTTGCGAC	60
CTTTACACTT GCTGTAAACC GTAACCTTAC GAATCAAAAT GGTGATCGTG AAGCAGACTT	120
TATCAACTGT GTGATTTGGC GTAAACCCGC AGAAACAATG GCTAATTATG CCCGTAAAGG	180

TACATTATTA	GGTGTGTGCG	GAAGAATTCA	AACTCGTAAC	TACGAGAACC	AACAAGGTCA	240
ACGTGTCTAC	GTAAGTGAAG	TGGTTTGTGA	AAACTTCCAA	TTGTTAGAAT	CTCGTTCTGC	300
TTCAGAACAA	AGAGGAACTG	GCGGCGGTAG	CTTTAATAAC	AACGAAAATG	GTTATCAATC	360
ACAAAATCGT	AGCTTTGGTA	ATAACAATGC	CAGTTCTGGA	TTAATAACA	ACAACAATAG	420
TTTCAATCCA	TCATCTTCTC	AGTCGCAAAA	CAATAACGGT	ATGCCTGATT	TCgATAAAGA	480
TTCTGATCCA	TTTGGTGGCT	CAGGTTTCATC	TATCGACATT	TCAGATGATG	ATTTACCATT	540
CTAAACGAAT	TTGATAGGAG	GGAACAAGAA	TGGCACAACA	AAGAAGAGGC	GGACGTAAAC	600
GTCGGAAAGT	CGACTATATC	GCAGCAAACC	ACATCGAATA	CATCGACTAT	AAAGATACTG	660
AATTGTTAAA	ACGATTCAAT	TCAGAACGTG	GCAAAATTTT	ACCACGTCGT	GTTACAGGTA	720
CTGGCGCTAA	AAACCAACGT	AAATTAACGA	TTGCTATTAA	ACGCGCACGT	ATCATGGGAT	780
TATTACCATT	CGTTAGTGAC	GAACAATAAT	TTTATGACAA	TTCCnGAAC	CAAAGAGCTT	840
GCTCTTTGGT	TCTTTTTTAT	CTTGACTAGT	GTTTCACGTG	AAACATAACA	AGGTTCGCAG	900
AATGAAGTGT	TTTGTGATAA	AATAAGGGAT	GAAATAACTA	GTGAATCACC	AGAGCAATAT	960
GCTCGTTGGA	ACAAGTCGCC	TCAAAGGTCT	GCAGAAAGAT	TGGAGGCAAA	AAAATGCAAA	1020
AGAAGAGAAT	TCAAAAAAAC	GGTTTCTTAA	TTGTTGTGGG	TCTTCTCTTA	GTAGAATTTT	1080
TCCTCTATTT	CTTACTAACA	AATAAATGGC	TGCTATTGGC	GGTAATTATC	GCATTAGATA	1140
TCTTTCTCTT	AGTGGTTATT	CGGCTGTTGA	TTAGAGATGT	AGAAATTACG	AACGTAGAAA	1200
AGATTCAAGA	AGCAAGTTCC	ATTGCTGAAC	AATCGTTGGA	TTATGTTGTA	AATGAAGTAC	1260
CTGTGGGAAT	TATTACGTAT	AACGGGGAAA	CACGCGCGGT	AGAATGGCTT	AATCCTTATG	1320
CTGCTTCTAT	TTTTAATAAA	GACAATCAGC	TAACGTTAAC	CGCTAGCCAA	GTTACGTCTT	1380
ATTTAGAATT	AGCAGAACGA	AACCAAGATA	TTTTTACGAT	TGACGAAAAT	ACCTATCGCT	1440
TTAGCGTCAA	TAAAGAACAA	CATACAATTA	CTTTTGAAGA	TATCACTAAA	GAAAGTAATT	1500
TGTATCAAGA	AAAAGTCGAA	ATGCAAACGG	CTATTGGCAT	TGTGTCTGTC	GATAATTATG	1560
ATGATGTCAC	CGATACAATG	GATGAGAAAG	AAATTTCTTA	TTTGAATAGT	TTCATTACGA	1620
CGATGGTTTC	TGATTGGATG	GACCAATACA	AAGTTTTTTA	TAAGCGAATC	AACGCAGAAC	1680
GTTATTTTTT	CATTGCCCAA	TGGGAAGATA	TTCAAAAAAT	GATGGACGAA	AAATTTTCTA	1740
TTTTGGATAC	GATCCGTAAG	GAATCAGCTA	ACCATGAAGT	AGCCATTACG	TTAAGTATGG	1800
GGATTGCTTA	TGGGGGCCCA	ACCTTAGATC	AAACCGGGAC	CACGGCTCAA	ACAAACCTAG	1860
ATACAGCTTT	AGTTCGTGGT	GGTGATCAAG	TGGTTGTAAA	AGAAGCCAAA	GATGAAGCGA	1920
AgCCGTTATT	TTTTGGTGGG	AAAACGGCAG	TAACGACGAA	ACGTTCCCAA	GTACGTTCTC	1980
GCGCAATGAG	CATGGCAATT	AAGGGAATTA	TTGCGGAATC	TGCTGACATC	TATATTATGG	2040
GCCATCGTTA	TCCAGATATG	GATGCGTTAG	GTTTCAGCATT	TGGTGTGCT	CGTTTAGCCT	2100
CGTTTAATAA	TCGAAAAGCG	TGGATTGTTT	TAGATGAAAA	TGAAATCATT	CCCGATGTCA	2160

AAAGAGTGTT AGAGGCGATT AAAGAGTACC CAGAATTAGA AGAGCGCATT ATTAGTCCTA	2220
AAGAGGCCAT GAAGCGCAAG AAAGAAAGTA GCTTATTAGT TATGGTAGAT TACCATAAAC	2280
CGTCTCTATC GATCTCACAA GAGCTCTATG AGCGTTTTGA TAAAGTAGTA ATCATTGATC	2340
ACCATCGACG AGGAGACGAA TTTCCAGCAA AACCCTTGCT TTCTTATATT GAATCTTCTG	2400
CCTCTTCTGC TTCAGAATTA GTCACAGAAT TGATCGAATA TCAAAGTAAT AGCGCAAATA	2460
AACTGCAGGC CTTTGAAGCA ACCATGATGT TGGCGGGAAT TGTGGTTGAT ACGAAAAGTT	2520
TCAATACACG AACGACGGCG CGAACATTG ATGTGGCTAG TTATTTACGA ACTTGTGGAG	2580
CAGACTCATC TTTAGTACAA TATCTATTAA GTTCTGATCT TACAAGCTAT CTGGAAATGA	2640
ACAATTTAAT CTCTAAAAGC GAATATGTCA CAAAAGATAC CGTCGTTGTT GCAGGGAGTG	2700
AAGACAAAAGA ATATGATAGT GTCACAGCTG CCAAACAGC GGACACATTA CTTTCTATGG	2760
CAGGGATTAA TGCAGC	2776

## (2) INFORMATION FOR SEQ ID NO: 142:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7337 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

TCCATTCCAG CTATTTCTTT CTACATCTGT AATATGTCTA ATGTTATCTT GATCATGCGC	60
ATCAAACCTCA GATTTAGTTG CCGTCTTTTG ATTTAGTACA TTTCCtwaAC CaACCTGakT	120
AGCTGTTACC GCatGaGGAT TACTTATATT TGATTTATGG GCATTAAAT CTTCAATATG	180
TGTAGCACTA TCTTGCAATT CTTTTAAAGC TCCGTGGGTT CTGCTAAAGA ACCAATTAAA	240
ATAATCAGCT GGTGGTTTTT GTGAAGCTTT CCAACCATCG CTAGTTAAGC TTTCGGGTGG	300
TTTAATACCT GGTGCTAACC AAACAGGCAA TTCCTTTGTA AATTTCAATT GCATTCTCCT	360
CTCTTTCTAC AGCGGCAGTT TATAATCATT TTCTGGAATA AAGACTCCTC CCAAAGTTCC	420
TCCGATATTA CCATCAATAT CAGCGAATCC TTCTTTACTA ATCTCTATAT CAGTCGTACT	480
TGAAAATGAA AAAGTTCCCT CCAAATTAAC GTAGGCTACT CTAATGCCTC CAGCACTGAT	540
ACTCTCGACA ATTTGTAAAA ATTGACTAAT ACTTAGACCC GAACTATTCA AGTAGTCAAG	600
AGGAGCTTTT TTTATGATTA CACAAGCAGG TTCTTTTTCA TCTACCGTCT CATTAGCTGA	660
TATAATATGG ATATCACTAG GATGACAATT TAAAGATGTC GCAATTGCAT GTAACATTTT	720
GTCGATTGAT CCATCGCTTG TAtTTCTAGC TACCTTTCCA CGAATCAATA CTCTATATAT	780
TTCATCAGTC GTTTTTcCTC TCGCTTGTC AACATTAGCA CCTAATTCAT CTAGTCCTTT	840
CCCTCGTGCT TCATCTATCG ATCGCCAGTT TGCTACTTTA TTCAGCAAAT CAGTCAAGTC	900
TTTAATTTCA AACTCAATAA TTTCAAGAAG TTTTGAAATA TTTGAATTTT CGCGATTGAA	960

TAGATCtGGT AAATAATCTC TAAGCTTTTT GGTCACTAGT AATCACCACA TTCTCTGTTT	1020
TACATTcAGC TGACTCATTG GGATTCAGAT TAATGTCTTT GACTTCGGTA GTTTCAGTAG	1080
ATAATCCAAT TTTTACATCA GCGACAACAA CGCCTGGnAT TTGATAAATT AATGGaTAAA	1140
TATAAGAAaAA TCTGACTACT TCACCCaTAG tCaGATkGTw AATGTAATTA TTTACTATAA	1200
CTTTTATTtC TTCTTGTCGG KTTTTTTCAA ATTTTGAATC TACTTGAATA GAAATATTCA	1260
CAAAAATAGG AACAGATTTT GCATAATCAA ATTTAACAAT ATGACTAAAT CCTCCTAAGT	1320
CTTTTACTTC AACTTCCTGA TTACCAACAG TGTCAATCCC AGCGGCAACA CTTTTAAAAA	1380
TTGCCTGCCC AATATCGTCT TTCACTCCAC CTAaAATATG CACATGTACC GACTTGGGTG	1440
GATTACCATA CGAATCAGTT TCCATTGTTT TATTTTCAAC AACACTAGCG GTCCGTACCC	1500
CACTAACTTC TAGCAAAGCA GTTAGAATTC CGTTTATTGG CGGTCCAGGA TTCCCACGAA	1560
CAGAAATACC AATTGGGTCG CGATAGGCTT TATCTGTTTC ACGTTCCTGCC CCACCCTCAG	1620
CCCTAGCAGG GTTATTAACA GATGATATTT CTTCACTAGG TTCCAATTGT ACTATAATAG	1680
TATTAGCTGG TACGTTTGAA CTAGCATTCT CTTCTAAAGA AATTGCGCGG CCTTTCCCAA	1740
ATCCATTATC ATCAATTTTG ACTATATCAA TCATCTGAAA CATAACTTTA TTTTCTGTTG	1800
AAAAACGTAC GCCTTCATTA ATGATGTAAC CAGGTTTTCC TGAGAACTCC AGCTCAACCA	1860
TTGCAACTGT AGCTGGATTT CTATAGATAC CACTATTGCG ACCTAAACGA TCCAATGAAA	1920
CCCCTGTAGC TTGACTTATG AAGCCACTAT AATAAACTCT TTCAGTTAGT TCATGAGATA	1980
TAGATAAAAA CCATGCAACG ATACGAATAA TGATACCCAG AACAGAATGT CCTGAAACAT	2040
TAACATCCGA ACCAAACAAG GCTTTTGCCCT TTTCAGACAT ATCATAGAGT ATTTTCATCAT	2100
ATGTTTTTCT TTAAATCCA TTTTCATCAA GCACGTATTA CCACCTCCAC TTCATCACCT	2160
AATGTTGATA GCATCTCCAC TGTAATATTC AACTGTCTAT TGTTTCTTAC AATTTCTATA	2220
TTTTCAATAC TATTAATTCT AGGTTCTTGA TCTAAAATCG CTTCTGTAAT ATCTTGTTTC	2280
AAGTAATCTT CATTGTAATT TTTTCCAAAC ATATTATCAC TTTCTAAACC AACGGATGTA	2340
TCTAATTTAA ATTCTTCTAG TCGTATTGAT AAGATCATGA ATACACTTTG TGCAAGTTCT	2400
AAATCTCCTT CAACTAGTAA AATTCCATAG TCAATAAATG ATAAGTCTCC ATTGACTATT	2460
TTCAAATCCC TCATTAAGCA ATCACCCECA TAATAACAGC GTCATTTTGA CTATGCATTC	2520
TATCAGATGA TAATGAATAA TCAGCAGAAC TTCTATAATT ATCTATGTCA CGATCACAAA	2580
AAACTACTAC TACAACCGCT CCCTGTGAGA TATCAGACTG ACAATGCTTA GTGATCAACG	2640
CGTTTAGTAT TAATGCTCTT TTGCCTCCAT CTGATTTTAA TGCCATTGGC TGAACATCAG	2700
CTCTTGACC TTTTACAGTG ACAACTCGAC ATAGTTGCAT AACATTAATT TCTTTTAAAA	2760
TTCTATTTTT AAATGATCGA AAAAAAGCTA AGTCAGTTTC TTTCAATTAGG TTACCACCTC	2820
ACACGTTGTA ACAAATCGAC TACCATCATA GGAGTGCTGG cCATTTTTTA CATAAAAATT	2880
TCCyTTTACA TTTTTtGAAT CGATATAGAC TGCTGTACCT GTAGTGATTT TATGTTGCAA	2940

AAGACACTCA ACAGACCAAC CAGTATAATC ATGACTTTCA ACTTTTGTTG GTTGGTTAAT	3000
AAGTCCCGTT TCAGGACTTA ATCTGTATCT CTCTTTATTT CCATCTCGAA AATTTTAAAT	3060
TACAAGTTGA CCTCTTTTAA AATATAGTGA TGTGTCACAA GCTTTGGAAA TCTCTTCTAA	3120
AACCATCATT GCTTGGCCAT CAGCTGTGTA GCCAGACCCA TATATTTTAT TATTTTAA	3180
CTTAATTTCT GACAAAGGAA TATTTGCTTC CCTAGCAACT CGATTtATAA TAGTAkGrGC	3240
ATCCGTyCCA TTATTAAAAG TAATATTTAC TTCTTTTtKGT yCCGAGTAAc TTTACCTyCT	3300
AAAAAAGTAA ATGTTGTTGC TCTATCTACc CGTTTAGAAC AGAGGGTTTT ACTTCTGCTA	3360
TCGTTCCCTGA CGTAATCACT CCATTTGAAG TACCTGCGTA TCCTGCGTGG ATATAGACTG	3420
GATTTCCTTT CTTAATAAAA TCAATACTTT TTTTATTTAA ATTGTATATT GTCACACTCA	3480
CTTCAGATGG ATTAGGTGAG TCAGAAAAAG GTGCTGTAAA GTGAATTTCT AACCGATCCA	3540
ACCTGCCTGA ATCAGCTCTT AGTAGAACTC GATTCTTCCC ATTCTTGTCG TGTATTTCAA	3600
TTTGTAATAA TCGTTGCCAT TGCGTATTTT CCATTTAATT ATCTTCCCTT TCCAAACTTG	3660
GAACATTATA ATTTCGGAGGT AAATCATCTA TATAAAGGAA AACCTGAATA CCAAAATTTT	3720
CAAAAGTTAT TTCTTTTGAA GCTCCCGACT CATCCATCGG CACCAAATCT GCTGAAGGTA	3780
AACGAGTATC TACAATGTCT TGCCAAAGTC TTTCATCGAT TACCATACGC TCACCAATAG	3840
CTATCGGTGT ATGATCTATG TCATATAGAT CTACAGTAAA AAATTTTCT GTTTGATTAT	3900
AGTCCACCTC AAAAATGTAA TTTGTATTAC CTAAAGGTAT TTCGAATTTT TCAGGTAATG	3960
AATATTTATC AATAGGAATA TATGCTCTTA AAGACATTTA GTCACCTCAT TTCACACGCG	4020
CACGAGCGCC TATGGGTATA AATCTATCTG GCCACTTGTT CCAATCTCTT AATTGCTGAA	4080
TAGACGTACC ATATTGTTGC CACCAACCCC AATAAGTATT TCCTGGCTGA ACTGTCACAT	4140
ATACCGCATT ACTTGTTTTC GAAGGTTGTT TTGCAACTGG CGGTTGTGGA ATCTTCTCCC	4200
AAATAGTCTT AGCTACACGT ATGGGTTGTA ACGTAATTGT TATAGTGAAA CCATTTTCCA	4260
CTGTGTCATA CCCTTTACTG ATATCTTGTA TAACTGCATT TTTAAAATAT GATCTGCCTC	4320
GATAACAAT CCAACGAACA TCTTTTGGC ATGCAACTAA TGTATTGTAA GCTTGCTCTG	4380
CAGCGTtACC ATTTTtagCG AGAATCCAGC CACTAATTGT GACTGGACCG CCTGTATACA	4440
TCATATTGTC AGTGATTGGT GCTCCTGaTT CAACAGGATA TTGAGATACA TTAGCAGCAC	4500
TGGTTACAGT CTCACTGaCG TTTACAATCT CAACTTTAGA TTTACCACTC TGaATGTATC	4560
CCATTAAACA ACCTCTCCTG TTCCTAGAAT ATTAAGTAAT TTTGCATACT CATCTTCTAA	4620
TGCTTGCTTT ACAGCCTCTT TAATATCACC AACACAGAT TTATCAGAAT TATTCCCAAC	4680
ATTGATAGTA ATATTAGGAC TAAAATTAAC AGATGGAGAT CTTGATTAA AGAGCGCTTT	4740
AGTTTTTTCA TGAGGATGTA CAGTTCCAGC TGTATCTGCT TCGAACAAAT CTGGTCCATT	4800
CTCTCCTACT AAAACTGTTT CACCTTTTGA AGGGCGGCCT CCCTTTGCAT AAGCCCTTAA	4860
TCTATGACCT GTTGGTCCCC ATCCAGAATG TCCATACGGC AAATCTGTTC TCCAATTAGA	4920

ATTATTGAAA AATGCCATTA GTTGATGGAA ACCATTTCATT ATATTTTCAT ATCCACGTAC	4980
TTTATATGCA TCAAAAGTTT GAGGAATGTA CTGAAGCAAT CCTCGAGCTG GGTTTCCACT	5040
AGCTGTATTT ACATCCCCAA CAGCAGAACT TTGAATAATA CTTTGTTTAC CACTAGACTC	5100
CCTCTGTATT TGAGCTAAAA TGCCGTTTAC TTCTGCATCA CTTACTTGTT GATTTCATTTG	5160
TTTTGCAGCT TTTCTAACTT GAGGACCCCA TCCTGCAGCA CCTACCGCCA CACCACCAGC	5220
AGCTCCGCCA TAAGTATCAG GATTTACGTG TTGGCCATTT GGTCCGCCTT TTCTAAGTtC	5280
ATAATGAACA TGAGGACCAG TTGACCAACC AGrAGrACCT ACATCACC GAATTGACC	5340
AGCTTTAACT TTATCGCCCA TCTTAGCTCT TATGCGGCTC ATATGTCCAT ACATAGcCCA	5400
TAGaTTATCm GcAACTTTaA TTCCTACGTG CTCACCTAAT CCAaTGGnAA GAAGATTGAA	5460
CCCaATCAAC CAAACCgGA TA\TTGAGCTG GaATAGGCGT cCAGTTGGAG CAGCATAATC	5520
AATACCGTGT GAAAATCCCC ATAGAGGCCT GGTCGCTTAC CATAATCAGA AGTACGAACA	5580
AATGGTGATC CAAATGTGG AGCAAAAGCA CCTGAGCCAA AAGAGACACC ATCTTCCGAA	5640
TCAAAAAAC CTTTGAAAGC TTCAAGTTTA TTTTTCACCC ATTCGGCACT GGTGTCTTTT	5700
AGTTTGTTCA TGACACCAtG TCCCAAACCT tGGATATTTT TACCCTTTTT ATAAGkGTTA	5760
TGcTTATTAA ATAAGcCAGT CACTTTTCCa ATCGGGTCTG ATAGCCAATC TTTTGcTGTT	5820
TCAGCTATAT CTTTAGTTTT TTCTATAGCT TTCGTTCCCTA CTTCTTTTGC TTTATCTACT	5880
GCATTAGATC CAAAGTCTTT TAATTTATCA ACAGTATTTG CACCGAAATC CTTAGCACTA	5940
TCAACAAAT CAGAAAGCCC ATTCCTTTA CCTTTATGGA AACCTGGTAA AACTGTTCTT	6000
TGCCCCATAC CACCATTTAA TACAGCTTTT GTATCTGAAT GGTTTAAAAT CCGTTCACCC	6060
GATGTTACAT TAGTAACTTC TGGACCATTT GATCCTAATA AACGCATTGA AGAAGTGCTC	6120
TTGTTATATG CTAATTCAAC ACCTTCTTCA CCAACTAGAG CTTGACCACT GTATGATGCA	6180
CCACGAGATC CTGTGCTAAA AGAGCGACCA TGCTTGTTTT TAGTAGAATT AGGTGTATTC	6240
GGTTTCCATT CTGGTATTTT AGGAATACTA AAGAAATCTA ATACGACATT AATACCACCT	6300
GTAACAGCAT TTACTTTATC GGCAGTCCCT GTTTTAAAAT TATCCCAAGT TTCTAATGCT	6360
CCAGACACAA CTTTATCCAT AGCTTCACCA AAGGCCTTAC CAAAACTTT AGCGCCTTCA	6420
ACTGTGCCGT CCCAAATTC TTTAAATTC TTTTGGAAAT CCTTTTAAA GACTTCCCAT	6480
TTAGATAATG TTTGTCCAGT TTCCAGTCC ACTTGTTCAA GATGACCTC CGCTTGcTT	6540
TTAGCCTGCG TTACAACATC TTCATGTTGT TTTTCTGCTT CTTTGACTAC ACCGTCTCTT	6600
TGTTCTTGGG CTTTTTTTAA TGCATCATCA TATTGTTGTT GAGTAATCGT ACCATTGACA	6660
TATCTTCTT CATCTAAAT GCGCTTTGTT TCTTCATATT TTTCATTTGC AGATTAAAT	6720
GTACCTTCTT TTGCTTTATA TGAAGCTGCT ACAGCATCAG CAGCCTGTTG CGCGCTAATC	6780
TCACCTGATG CATTTTTCAA ATTACCATA ATGATTTTTT GCTCTTTAGC TGAATCTGAT	6840
AATGCAACAA CAGCTTGATT TTTCATATCT TCATGAATCG ACTTATTAGC AGCAGCATGT	6900

840

TTTTCTTCAA CAGCACGTAC TGCAGCAGCT GAGGTTTTTT CTAACGACTC AATTTCTTTT	6960
TTATCGTTTT CTGATAGTTC TCGATTTTCT AATCTTGCTT TTTCTTTGAT TTCATTGATA	7020
CGATTTGTGT AATACTCTGT AGCAGTAATA GATTCATCAT ATTCTTGTTT TTCTAATTC	7080
TTGAAATCTT GAACTTTTTT AGAGAACATA TTTGTTCGTA CTTTGTCTAA TTCATCAGCA	7140
GCTGCCTTGG CTCCTTGTAAC ATCCTTTtCA TTAAGAATTC CCATACCAGC TAATTTATCA	7200
AGATTGCCAT TAGATTTATC TkCTkTTTA TkCAGCTGTT TAGTAACTTG ATTCTGCATG	7260
TCTTCTAAAG CTTTATATCG TTTAGTAACT TCTTCCATTG GTAAGAACTC GTCCTTCAAT	7320
TTTACCTGTA GTGTAnT	7337

## (2) INFORMATION FOR SEQ ID NO: 143:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1160 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

ATAAGGAGGA AATCTGATAT ATGAAAACAA AACGCTTTCA GCGCACGGCG GCGATGCTGC	60
TGTGCATTTT TATCTGCGCG TGGACAATGT TCCCCCTGAT GGGTGCGGCC TTGCGkCsG	120
AGAATGGGCT GCATGACACC GCTGAACCGG AAATCACCTG CTTCTGTGAT ATTCACTGCA	180
CGGCGGATGC CCTCAATGCA GACTGTCCGG TTTGTACACA GGATGTGAAC GCCTGCACCG	240
GCACGGAACC GGAAGCGCCC ACCGAAACAG AAACACCCAC CGAGCCGGAA ATCACCTGCT	300
CCTGCAATAT CCGCTGCACG GCAGATACCC TCAACGCAGA CTGCCCCGTT TGTACGCAGG	360
ATGTAGCTGC CTGCACCGGC GCAGAGCCGG AAATTGCCTG TTTCTGCGAT ATTCGCTGCA	420
CGGCGGACAC CCTCAACGCA GACTGCCCCA TCTGTGTGCA GGAGGCAGCC GCCTGCATTG	480
GCGCGGACCG GAACCCGTCT GTACTTGTGA AACGCGGTGT GCGGCAGACG CACCTGACAC	540
GGATTGTCCC ATCTGTACGC AGGATGTAAC CGGCTGTACC GGCAAAGAGC CGGAACCGCC	600
TGCCAAACTG GTCTGTATCT GCGCCGCCCA CTGTGAACCC GCGCGGTCA ACGCCGACTG	660
CCCCGCCTGC CTGTTGAACC TGGCAGACTG TATTGTGGAA GCCCCACGC CCGTCCCGGT	720
CTGCGTCTGC GAAAACCGCT GTGAACTTGG TGCGGTCAAC GCCGACTGCC CGATcTGCAG	780
AACCGACCTG GCCGGATGCA CCGGcAAAGC CCCCcGCCA ATyCCTGCAT CCAATCTCTC	840
CATCAAAATT ACACCGCCCT CCGGCTGGGC AACCGGCAGT GCTGCCGCCG AGTTTCGCAT	900
TACCGATGAA GCCGGGAACG GCTTTGCGCT GGTGCAGGTG AAAATCGAGA AAAACGGGCA	960
ATGGCGGGAT GTGACGGACA GCTTGAAGCA GAACGAAAAC CGCTGGTATG GGGAAATCGA	1020
CCTTTCCGGAG AACTGCACCG TCTATGTCTG CGCcAACGGC CATGATGGGA AgTGTATGAG	1080
AAtCCCGCTA TATTGAGTGT TTCGACCGmA CggcCTGCTT ACGCGcCGGC ATkGAcGGcG	1140
GtTsTcCGTG CGGAACAAGC	1160



## (2) INFORMATION FOR SEQ ID NO: 144:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12860 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

TTTACnTGAA CCAATTTCCA ATGTGGATGG GATGATTAnG CCAAACGTTC TGGCCATTnC	60
CCTGGCCATT TTTCAGGACC TTCTACCACA TATTGTGGAT ATTTTGCCAT ATATTCCGCT	120
GCTCCGGGTC TATTCTTAC ACGCATATTT AAAATTCCTT TCTTTTAAAA ACAGACTGGT	180
GAAATTCGCC AGCCTGTGTG GATGTTAAAT GAATTAGTTT AAATACTAAA CATTTCGTTTT	240
AATTGTAAAA TTTCTGCAGT CATTGCTCTT CTTTCGCCCT CTTGGTGATG ACGAACAATC	300
TCTTGTAATA AGCTAAATAA ACCATACCAA TAAATTCGTT GAACCGTACT TTCATTCCGG	360
ATTAAACCAT AAGATAGTAA CCATTGATTC CAACCGGAAC GAGGGACATA ATGACTCAAA	420
ATCATTCCTA AATCTACCGC TGGGTCCGCG ACCATCACAG AATCCCAATC TACTAAATAG	480
AGATAATTAT TCGAGACTAA CCAATTCCGA TGATTCACAT CGCCGTGAAC AGCTACATAA	540
CTAGTTTCTG GATATTCTGG TAAATGATTA TATAAATAAG TGATTACTTC TTTTAAATAA	600
CTGTTTCGTT TTAATTCCTT AGGTAGTTGA CGCTTATACT CTTTCAACAT TGTTAGCGGG	660
GTACAACTT CGCCACCAAT TTTTTTCAAC ATGTCTTTCA GCAAATTGGA ATGATGCAAT	720
TGGTATAACA CGTCAATCAC GTCATTGCGT TTGCCAATTT CGTCCGCATG TAATAAACGT	780
CCTTCAGCC ATTCTTGGGC TGTTAGTACG TCGCCATTCG CTGTCCGCTT TGTCCAAATC	840
ATCTTCGGGG CTAACCCCTC TTTAGACAAT GCTGCAAGAA GTGGAGAAGT GTTTCGTTTA	900
ATGAACACTT yCTCGGTTTC TtkTATTCCC ATGTATGTTT GACCAGTTGC GCCCTTTATC	960
GGCTGCAGAC GCCAGTCTTT ATCCAAGTGA AATTCATGG AAATTATCTC CTATCTCAAC	1020
ACAATTTTAT TGCAATTTTT ACAGCAATTA CCATTTTAGT CGCCAAACCT TGCTTCGTCA	1080
AGCTTCTATG TGGTTCTTTT GTAAATTTTG CTGGGTTTcT CATCTTTATA ACTTATTCAA	1140
GCGCATTGGA ACATACATTT TAGTGAAAT CACAATTCT ACAATTAAGG CACCCAGAAT	1200
GATCAAGCTA TTTTGCAAAC TGCCTAAACG AAGGCAGACA ATCCCACCAA AAATAATGGC	1260
CGCAACAGAT AATGCGCCGC TAATTAACCG GCGAATCGCT GCTTGTTTTT TTTCTATTGA	1320
AATTGGATAT AGCTGTGTCA TCACCATATA ATCAAATTCG GTGTACATGG GAATCAACTG	1380
GAACCCGATT AAATAGACAA ACAATAAAGC AACTACCAAC GCAATCCAGA ATTCTTGTA	1440
AAAGAAGAGT AGCACGCCAC CAACCAAAAT CAAACGAAAC AATAAGCCGC TATATTCGCT	1500
ACCTCGTAAA AAGCTCCGAA TGAATAAGTA AGCATACGTG TTCTCAGATG TTTTTTTGAC	1560
TACTCCGAGG AGAGGATCTA AGTATTTACG GCGTTTCACT GTACTACTCA CTTCAGGAAT	1620

ATCTGTAAAC AGCTGAATGA ATTTATAAAT TCTATGCATT CTGTTTTGTT CTTTTTGGAT	1680
CATTTTTTICC .CAATCCAAGA AACTCTGTTT CATCTTATTT TGCCAAATAC TCGATAAACA	1740
AACGGCCAGA ATGATGCCAC CAATCACACC AACCCAAGGC GTCCAGTAAA TAGCTCCCGT	1800
CATAATCAAG AGACTACTGA TCAACCAAAC AAACCACCAA CTACGGTAAG TGGTTGACGA	1860
AAGATGATAG AGCCCATAGC TTTGTAAACT AAGATGTGTA TAAATCATAC AAAC TAGCAT	1920
GACCAAAAAT AAAAAGAAGG TTTGAAACGA CCAACCTGTT GTGACTACAA TCAACGGCAT	1980
CAATAGGCCA CTTCTTAJAA AGCTGACTAC AATCGGTAAA AGTAAAGAGT AGCGCAGCGC	2040
TCGTTTCAGA TAAGCAGCGA ATTGTTTTTC TTTTGGTAAT AAAAACACCT TGTCTGGCTC	2100
TTCTGTCAGA GTAGCCAACT TACCAACTTG AATGAGTAGC AGCCAGCCCA ATAGAATTAA	2160
CGGTCTGCCC CAAACAAAAT CACGAGGTAA TTCTTTTAAC AACTGGGAAT AGTACAGACC	2220
TAATCCGCCT AATAAAAACA TACAGACTAA AACAAAATGA TCATTTAAAA TATAGCGCAA	2280
GTACTTGGAC ATTTTTTTGA AATGCCGTGA CAAACGTTGA CTAAAAAATT CGCCCATGTT	2340
ACCCACCTT TTCTTCCTTC GTCAAGGCAA GATAAATATC ATCTAAAGAA GACTCTGGTA	2400
AATTAAATTC TGCACGTAAT TCAGCCATTG TACCGTTTGC CCGTAATTC CCTTCGTGTA	2460
GCACCACAAA ACGATCACAA TATTTTTCAG CTGTTGCTAA AATATGCGTG GACATTAAAA	2520
TCGCTGCCCC TTGCTTACGC ATCGTATCCA TTAATTCTAA TAAAGCATGA ATTGCTAAAG	2580
GATCTAAGCC TAGAAAAGGT TCATCAATAA TATATAAACT CGGTTCAATT AAGAATGCAC	2640
AAAGTACCAT GACTTTTTGT TTCATGCCTT TTGAAAAATT AGCAGGAAAC CATTCCAATT	2700
TATTGTCTAA ACGGAACGTT TTTAGTAATG GTTCTGCTCG TTTGAACGCT TCTTCTAATG	2760
GAATATCATA AGCCAAGGCC GTTACTTCGA TATGCTCTtT CAATGTTAAT TCTTCATATA	2820
AAGAAGGTGT TTCCGGAATA TACCCGATTT TTTTACGATA TTCTTCAGGG GCTTGTTGCA	2880
ATGTTTCCCC GTCAATCATA ATTTTGCCTT TTTGCGGCGT AAGTAACCCA ATAATATTTT	2940
TGATGGTGGT ACTTTTCCCC GCACCATTCA AGCCAATTAA ACCAACCATT TCACCAGACT	3000
TGACGTCAA AATTAATATCT TTTAAGACAG GAATATGGCC GTAACCTCCT GTTAAATGTT	3060
CAATTGTAA GCTCATTTCT TTTGCCTCCA ACCTTTTCTC TCTACCATTA TTAAACCCTA	3120
AGCGAACCAA GAATTCAAAA ATTACTCCCT TAATTAAATG GATTTTTTTC TTTCTCTTAT	3180
TATACACGAA ATGAAACGAA CTTTTTAGCG AACTCACGTA CTTTTTCATC CaCAGAGAAA	3240
TCAACTTGTG GTAAAGTAAA GGAAAGAACC TAGAGAAAAG AGGAATTTTT CATGAATGAT	3300
TGTATTTTTT GTAAATATAT CAACGGCGAA ATCCCAAGTT ATAAAGTCTA TGAAGACGAG	3360
AAAGTCTACG CGTTTTTAGA CATTACACAA GTGACCAAAG GGCATACCTT AATGATTCCA	3420
AAACAACATG TTGCTGATAT TTTTGAATAT AATGACGTTT TAGCAAGCGA CGTCTTTGCT	3480
CGTATTCCTA AAGTAGCACG AGCATTAGAA AAAGCTTTTC CAGAAATGGA AGGATTAAAC	3540
ATTTTAAATA ACAATAAGGA AGTGGCTTAT CAATCCGTTT TCCATTCCCA CGTTCATTG	3600

ATTCCCGTT ATTCAAAGA AGATGATTTC TCTATTCATT TTGGCAACCA TCAAGAGGAC	3660
TATTCGCCG AAGCGATGCA AGAAATTGCT GAGACCATTG CAAAGCAGGT GAACTAAATG	3720
AAAAAATTTA TTAAAGGTCT CTTTTTTGGC GCAgCTGCTG GTACCATTGG CGGCTTGTTG	3780
GCTGCTCCAA GAAGCGGGAA AGAAACACGC CAACATTTAA TCAATGAGCT AGAGGACTAT	3840
CGTTCACCTGA AAAATCAAGT CACGAATGAC TGGGATCAGG TGCAACGAAA TTTGGCAGTT	3900
GTCGAAGAAA ACGTTCCTACT GGCAACTGAA TTTTCAAAAG ACCTGCAACA AGAAATCACT	3960
GATTTCAAAT TCCAAGCGGA ACCTCGGATT GCTCAAATTA AAGAACAAAT TGCTAAAATC	4020
ACGGCTGAAT TACCTGACAC ACAAACCAAC AAGCAAAAGT AAAGTCTTTT ATGTTAATTA	4080
TTTCTTTTGT TATTAaaaaa ATAAAAAGTG TGACGTAAAT ACGAATTTTG TTTGAAACTA	4140
TGCTGAGAC TCTTTCTTTT TCAAAATGAG GTAtGGTATA GTTATAACAG AwATAAAACT	4200
ArAAAAACA GGAGTGCATA AGAGAATGAa GAaAAAACTA ATCTTAGCTG CAGCGGGCGC	4260
AATGGCCGTT TTTAGTTTAG CAGCGTGTTT AAGCGGTTCA AAAGATATCG CAACAATGAA	4320
AGGTTCAACA ATTACTGTTG ATGATTTTAA TAACCAAATT AAAGAACAAA GCACTAGCCA	4380
ACAAGCGTTT AGCCAAATGG TTATTTATAA AGTCTTTGAA GAAAAATATG GCGACAAAGT	4440
AACTGACAAA GAYaTTCAAA AAAACTTTGA CGAAGCCAAA GAACAAGTAG AAGCACAAGG	4500
CGGAAAGTTC TCTGATGCAT TAAACAAAGC TGGTTTAACT GAAAAACAT TCAAGAAACA	4560
GTAAAAACAA AGAGCAGCCT aTGATGCAGG TCTAAAGCC CACTTAAAAA TTACAGATGA	4620
AGACTTAAAA ACAGCTTGGG CAAGTTTCCA TCCAGAAGTA GAAGCACAAA TTATCCaAGT	4680
TGCTTCAGAA GATGATGCCA AAGCTGTCAA GAAAGAAATC ACTGACGGCG GCGATTTAC	4740
AAAAATTGCT AAAGAAAAAT CAACAGATAC TGCTACGAAA AAAGATGGCG GTAAAATTAA	4800
ATTTGATTCA CAAGCAACAA CTGTTCCTGC CGAAGTTAAA GAAGCTGCCT TCAAATTAAA	4860
AGATGGCGAA GTGTCAGAAC CAATTGCTGC AACAAATATG CAAACCTACC AAACAACCTA	4920
CTATGTAGTG AAAATGACGA AAAACAAAGC AAAAGGCAAT GACATGAAAC CTTATGAAAA	4980
AGAGATCAAG AAAATTGCTG AAGAAACAAA ATTAGCCGAT CAAACATTTG TTTCGAAAGT	5040
CATTAGTGAC GAATTAaaaG CGGCCAATGT GAAAATTAAA GATGATGCCT TCAAGAACGC	5100
TTTAGCAGGC TACATGCAAA CTGAATCTTC AAGCGCTTCT TCAGAGAAAA AAGAATCAAA	5160
ATCAAGTGAT TCTAAACAA GCGATACCA AACAAAGCGAC TCTGAAAAAG CAACAGATTC	5220
TTCAAGCAAA ACAACAGAAT CTTCTTCTAA ATAAGCATAA AAAAGAGAGT GGTGATTGA	5280
CCACTCCCTT TTTTATTTA TCTTTTGGA CATAGAACT GCGGTATCT AAACCAAGA	5340
TTGTTCCGT ATACTCCCCa GGCTGTGTTT GACGAATTGC CCCTAGCATt CATTGAATG	5400
GACGCATCCA TGGTATCAAT TTGATGTAAG ATTCCGCTT CCATAATTCT TGGTCGCACA	5460
GGTGAGCCAT ATTCTAACAA TCCATGGTGG GCTAAGACCA TGTGCCGAAG AACGACAACA	5520
TCTTCTTCTC GATCATCAAT TTTCAAAGCC AAACACGCCT TGGTAATTTT CTCATCCACC	5580

AAAAC <sup>1</sup> TAAAT GACCAATTAA GTTCCCTGCC ACCGTATATT CTGTCGTAAT AGCACCAGAT	5640
AGTTCTAAAA CTTTTCCTAA GTCATGTAAA ATAATCCCCG CGTATAATAG TGACGTGTTA	5700
ATTTCA <sup>2</sup> GTAT ATTCCTTAGC AATTGCTTTG CCTAGGCGCA ACATTGAGAC AGTGTGGTAA	5760
GCTAAGCCAT TCGCAAAAGC ATGATGGTTC CGTTTGGCTG CAGGAAATTC ATAAAATTCT	5820
TGTTGGTATT TCGTCATGAG ATAGCGCACA ATACGATTCC AATGAGCATT GGTAATTTCA	5880
AAAATCACTT GATTGATTTC CTCGACCATC TCTTCTTTTT TCAAAGGCGC TCGTTCCATA	5940
TATAA <sup>3</sup> CTTG GCTGGCTTGG TTCCTCTGCT GTCGCTAAAC GCATGTGTAA AATTTTACT	6000
TGGGGTTTAT TTTGATAAAC TTCTCGtTTC CCATTCA <sup>4</sup> GTA AAACAACGTT TCCCGCAGTA	6060
AATCGTGTAA TTTCA <sup>5</sup> TCTC AGAGGCGTCC CAATATTTtC CGTCAATCGT GCCAGATGTG	6120
TCTTGAAAGG TAAAGGCAAT AAAC <sup>6</sup> TTTTTC CCATTTT <sup>7</sup> TAG CTACACGTAC ATCGGCATTT	6180
TTAATTAGAA CAAATGATTC AAAAAGTTCT TCTACTGTTA ATCACGTAA TTTTTC <sup>8</sup> ACT	6240
TGACGCTTCC TCCTTCTAAC GCAATTACTC TTTGTTGTAA GTCCTGATAA TAGCTTACCA	6300
TTTCTTGGTC AGAAGAAAAA CAAATGACTT GCTGTTTTTG ACCAAATTTA GTAAATAATT	6360
CAGCTAATTG GGCTTTTCTT TGATGATCAT AGTGTAA <sup>9</sup> CCA GCCATCATCA ATGATAATTG	6420
GACAAATCAG TTGCTCGCCT TGGACTGCTA AAAAGGCAAA ACGAACAGCC ATCATCAATT	6480
GATCTTTAGT GCCAGTTGAT AATTCATGTA AATAAAACAT TTCTTGTTCT TTCGTCACCG	6540
CAACTAGTTG ACCTTCGGCT ACTTG <sup>10</sup> GATTG ATTGATAACG CTGATTGT <sup>11</sup> T AAAAGTGCAA	6600
AGTAGCTTGA TGCATACTGT AATAAGCTGG GCAATTGTTG TTCCGAAAGC TCTGTTAATA	6660
GGTCCATTAG CAGTTGTCCC GCTAGTTGAT ACCCAGACCA ACGTTGAGCT AGTTCTTTGA	6720
TTTCCGCTTT TAA <sup>12</sup> AATTGCT TGTCGTTGAT ACAACTCATC TAAGGTACCA TCCGTCATTA	6780
ATTGTTGTTT TTCATAAATC AGCGCTGAT ACTGTTCTTG TATTTTATCT AACTGGCTTT	6840
CTTCCAATGC TTGTTGctGC GCCAGTTGTA AAGAACGGTC TTTTAACGTT GCTTCTGTTA	6900
CCGCTTCTGG AAACAGATTA CCTAATTGCG TCGCTAATTC TTGCCACGA CCAACTTGGG	6960
CTTGCTGAGT GGTCAGCTGT TGAATCCTAC TTGGAACATC AGCGATTGAT AGAATCCGAT	7020
AACGAATTAG TAAAGGCTGA ATTCGTGTTA ATGCTTCTTG CTGTTTAGTT TTCAACTCTC	7080
GTA <sup>13</sup> ACTGCTG TTTGAGATAG CCATCTGCTT GATACTCTTG AGCAAAGCGA ACTTTCTCCA	7140
TTTCTTCAGC AAACCCTTCT AAAAAGTGTA ACTTTTCGGA AAGGCTCTTT CCTTGCAAGG	7200
GTAAATGCGT CATCAATGGG AAGATTGT <sup>14</sup> T CTTCAAATAC TTGGCGTTTT TGTTGATTTT	7260
CCTTCAATTG ACTCATGAGT TCTTCATTTG TTTGAACAAG CAATAAATAG CGCGTAATCA	7320
ATCCCCGCTG ATTCATTAAA GTATCAATTT GTTGCA <sup>15</sup> TTTT TCCTAATTGA TGGACTTGGG	7380
CTTGCTCCGC TACTGCTTGT TCAAGTTGTT CCACTTGCAG TTGTGCCTCA CGTTCTTGAT	7440
TCAGCGCCTG TTGGA <sup>16</sup> ACTGT TCATTTAAAT AATCTAGTTG CGAAAGTTTT GTCTGCCACT	7500
CTTCTTTCAC TTCATCAGAG GAAGCTTCTT TTCTAGTGGC GTAgcTAAGC ACTTCTTTCG	7560

CGATAAACT AACTCCCACG AGAACTAGGC CCCATTTTAA AGGGCTTGGC AAAAAGAAAC	7620
CGAGTAATAC AAGGATACTT CCTAGCAATA ACCAAAGTGG TTGACTTGTT TTCTGTGCTT	7680
TTTGGGCACG CGTCGCTTCA CCAAAAAATG CGGCATGCGC TTTTCAAAA TCTGTAACT	7740
GGTCTTCCAG ATTTTCAATT TGAGAAGCCA AGAATGCTTT CTGTTCTTGT CCTTTTGA	7800
GAGCTTGCGT ATACGCAACT TGTGTGTTGGC GCCACTTTTG AACGGTTTCG TTGTCATAAA	7860
AAAGCTGTGG TGGTCGTTCA GGCGACCATG ACCATTCGTT CTCTAAAAGG ACCATTCTT	7920
GACGATTTTG CTCAAAGGTT TGGGTCATCC ATTCCGATTC CGTCATCAAT TGTTGCATAT	7980
CATAGCGTTG ATTCAACAT TGTGAATCA ATGCTTCTTG TTCTAGGTAA AAAGCATATC	8040
CTTCTGGAAC ATCGAGCGCT TCTGACTGAA GAGATAACTC TTTGGTCAAA CGTTGTCGTT	8100
CTTTGTCTAG GTACATATAT TCTTGAAgG TTTCTTCTAA TTGaACTTGG TCTTCTTTA	8160
CTAAAAAAC GTTCTCTGTT TCTTcAGGTA GCGACTGCCA TTCTTCGTAC aAGGGAAAGT	8220
GaCGTTgcTg TTCCgCAACC TGTAACAAGT cCTGTTGCGT TTCTTTTAAT TGCTGTTGCA	8280
GTGtCACTCG CTGTTGTTCT GTTTCGCGAA TCGTTTCTTC TAATTGTTGA AAAGTTTGT	8340
GcTGCGCTTC TTTCTCTGA ATTTGCTCTT TTAGTTCTTG ATAAGCCGCC AACTTCTGAT	8400
TGAGTAATGG CTGCTGTCCT TTACCTTTAA AGATTTTTTG TGCTTCTTTA AAGTACGCAT	8460
TTCGGTAAGT CAAAGTTGT TGA CTACCAG AAACACCAAT GGCTAACAAG GATGTTTGCA	8520
GTTCTCTTC TGTCAACTTG TCACTCGTAA TTAATTGTTT TTGTTGGAAA GTAAAAACCT	8580
CTTGAAATAA CTTTTTCGTT AGCGGATGTA ACATTTGCTG TAAGGTTTTT TCATCCCCCA	8640
TTGTTGTTG ATAATAAATG ATTGCTTGCC CTTTATTTTT TTCTTTAAAA CGCTCGACTT	8700
GAACTTCCCC AAACACAGGA TGTGCCAACC ACAAACGACC ACCATACGAG CCACCATTTT	8760
TAGGTGCGTA ATCGCGTTTT CTTTGGCCCT TTGTGGGAAA CCCAAACAGC ATTGCTTGAA	8820
TAAATTGATA CAATGTGGAC TTGCCAACTT CATTTGCGCC GTACAGCAAT TGATTACCAG	8880
ATTCAAATGT AATTTTCTGT TGTGTCATT TTCCAAAGCC CACAATCTCT ACTGCGAGTA	8940
ATTCATAGT GGGTCCTCCT GAATAGTAAA GTCTTCATTA ATTTGCTCAT GTGCTAGGCG	9000
CAAGCTCTCT AATCGCCACT GCTCATCCAT TGAAAAAATT CCAGCAAATT CTGGATAACG	9060
TAACAATTCC TTGGTGGTAT CTTCAAAAAT AACTGGATCT AAGTATGTTT TTTCTAATGC	9120
CGTTAATAAG TTGGGACTAG CGGCTAAATA GATCGGTTCT TTCGTTGGTT CCTCTGCTTG	9180
CAGTATTAGG CGATAAACAA AAAAATCACC TTGCGTTTTT TGTAAGGCT GCTGCTGTAA	9240
GTAGCTTAAT AGTTCACCAT TGACAATGGC TTGCACCACG GTCACATCCA AATGTTGCGT	9300
TTCTTTCAGT GTCAACGCCA TCAATTGCTG TTGCTGATGC GCCTGCCACT TCGTCAGCAA	9360
ACTGGTTTCA AGATAAGACA AAACGCTTTG GGTGTCACGG CATTGCGCAA GAGATACTTC	9420
TTCATTCGTC CAATCTATTT CTGCGACTTT TTCAAATTGA ACAGTCGCTT GATTTTGGGA	9480
AAGTGTGACC ACTGCAACCC CTTGTAAAGA CTGTTCTTTT TTCGTGTGTC CTTGAGGTGT	9540

TCCAGGATAA ACAATTAGCG GTTGTTCACT AACAAATTGC GGCTGATGAA TGTGGCCCAA	9600
TGCCCCAATAA TTATAGCCAG TGGCTTTTAA ATCCTTCCAT GTAAACGGAG CATAATTTTG	9660
TTGCGCTTTC CGACTGGTGT CTCCATGATA AATCCCAATA TGAATATCAG ATTCTGCTTG	9720
TTTGATAGGA AATTCTAGTG CTTTATTTTC ATCAATCCAG GGATGCTCAT AACTAAACCC	9780
TGAAACAGCC ACTTGTTTCGC CCGCTTTCGT TTCAAAATAC AGTGTCTCAA CCATTTCTTT	9840
TTCAAATAAA AAGACATTGT CTGGAAAAGA AAACCAATAG CGGTCTTTTT TATAGTAATC	9900
ATGATTCCCA AATGTTAGAA TCACAGGGAT TTCTGCTTGT TTCAACTGTT CTAATGCAGC	9960
CATCAATTGG GCTTGCATTG AAATAGAGGT CTGACTTTGA TGAAACGTAT CTCCTGCAAA	10020
AATAACCAAG TCCACCGCTT CTCGAATGGC AACTGTTACG ATATTCATTA ACAATTGTTG	10080
ATTAGCTTCT CTTAaGCGcT TCGCTAAGGC TGTGGGATC CCAGTTAACC CTTCAAAAGA	10140
CCGATCCATG TGTAATCAG CTGTATGCAA TAGTTTCATC GTTCCCTC CTTTTCTCCT	10200
TCCTTTATCA TATCATTTTC TGCCAGCTTG GTCAGTAAAA ACATACGAAA ATTTGTTTCGT	10260
TTTTTTATTT TCAGTAAAG AAAAATCTG ACAAATTGTC TTATAGACAA CTTGTCAGAG	10320
TTCAATTACG TAAATTAATC GTTATACATA TCACGCACAG GAGTCATGAT AATTCGGTTT	10380
AAATCATTGA CAACTAAGCT AAACGCTTGT TCTTTTTCGCA TTAATTCGTT GATAACTGCT	10440
GTTTCTTGCA CTTTCGTTGC AATGCTTTGT GCTTTTTCAG CATCTTCATC CGTGAATTCT	10500
TCGCCGCGCA TTTGTTTTTC TTGTAGTCTT TGTTGGAATC CTTGGAATTC TTTAAACAAT	10560
GTGTAAGCTG CTTTCGCTGC TTTTACTTCT TCATACGCTG CTTGTAATGC TTTAAATTCTG	10620
TCTAACTCAC GTAGTTCACG TTCAATCTGA TTGGCTGTAT CATAAATGTT ACTCAATGGA	10680
ATCCCTCTTT TCTCAATAGT TTCATCTTTA TTGTACCATT TCTCTTAGGT TTTCCCACT	10740
AATTTCTTAA TAAGCCTCCG AAAATTCCTT CAATCTTGTC TTTGACTTTT TTCGCGCCGT	10800
CTTTAATTTG TCCCCAACT TCATTCAATG TATCTCCGAA ATTGCTAGCG CCGTCTTGAA	10860
CAGAATTTCC AAAGCTTTC AAGTTATCCC GCCAATTAGA ATTGGCACCA CCAGTTAAAC	10920
TATCGTTGCT ATCAACAAC TCCACCTG TGGCATAAGC ATCCGCGACA GGAAATTGTG	10980
CTTGTTTGAC TTGTGGCAAG ATGCCACTGG CTTGACTGTT GAACACCTGT GAGGCATACG	11040
TAGCGCTACT GCCTTCCAAA TAATGGGTTT TACTGGTTTC TTGGAAGCCT AACCAAGTAG	11100
CAATGACAAC TTCTGGTGTA TAACCAATCA CCCACTGATC GTTCGTTTTA CTACTATCGA	11160
AGTTGGTTTC TGTCGTACCT GTTTTACCAG CCATTACGTA GCCTGCTGGT TGTGCGTTGA	11220
CCCCGCTCCC AGAACTAAAG ACCCCTAGTA ACATACTTGT AATCCCGTCT GCAACATCTT	11280
TGGTAATGAC TTGTTCTTTC TTGACTTTAG TATTATCAAC TACAATTGCG CCAGTTGAGT	11340
CGACGATTTT TGTAATTAAA TGCGTTTCAG TTTTATAGCC TTCATTGGCA AAGGCACTAT	11400
AAGCTCCCGC CATAGTTAGT GGCGAAACAC CCGTTTGTA TCCTCCAAGG GCTAAACCAT	11460
AGTACCGATC TTTTCTGAT AACGGAATGC CAAATTTCTC TGTTTTTCA TAGCCTTTAT	11520

CTAAACCAAT TTGATGTAAT AACCATACCG CTGGTAAATT TAAGCTTTCT CCTAAGGCTT	11580
GATACATAGG AACTTCTCCA CTGTACGTGC GGCTATAGTT TTGCGCTGGA TAATAATCTT	11640
GTGGTTTATC TTCTAAAACA GAGTCTGGTT TATAGCCAGC CTCTAAAGCT GGTGTATAGA	11700
CAGAAATTGG TTTTAAGGAC GAGCCTGGTG AACGTTTTGT TTGCGTTGCA AAGTTAAAGC	11760
CTCGATAAAC GTGTTCCCTT CGACCGCCCA CCAGTGCTTG CACGCCGCCG GTTTTCGGAT	11820
CAATGGCTAC AGAGCCACTT TGAACCATGG CACCGTCTTC TGCATTGGGT GGAAATAGTG	11880
CATCATTTTT ATAAGTTGCA TCCATCGCAT CTTGATACCC CTGATTTAAA GAAGTGTAAG	11940
TCTTGATCC TTTATTTAAA ATATCTTCTT CTTTAAATTT ATAACGATTG ACCGCTTCAT	12000
CAATTACCGC ATCAAAATAA TATGGATATT TGTAAGTATT TTCGTCGCCA ACATAGGTGT	12060
CATTTAATAG ACTCGCTAAG TTAACACTCG CTTCTTGATT GGCTTCTTCT TGAGATAGTT	12120
TTTTATTGTC GACCATTAGT TGTAACCGG TATTCCGTCG CGCTGTGGCA TTGTCAATGT	12180
AATCAATTGG GTTATAAATT CCTGGCCCTT TTAACATGCC CGCTAGCGTT GCGGCTTCCC	12240
CCACTGTCAC TTGGGACGCA TCCACGCCAA AGTATTTACG AGCTGCATCT TGCACGCCCC	12300
AGACACCATT ACCAAAATAA GAATTATTTA AATACATTGC CAAGATATCT TTTTTCGAAT	12360
ATTTCTTTTC GATTTCATC GCTAAGAATA ATTCTCTGGC TTTCTATCT AACGTTTGT	12420
CTTGCGTTAA GTAAGCATTT TTGGCTAATT GCTGCGTAAT TGTACTCCG CCACCACCGC	12480
CACCGATTTT CCCGAAGGTC AATTTTCCAA CAACCGCAGC GGCAATCCCT TTGATGGAAT	12540
ACCCTTTATG TTGATAAAC CCGCGATCTT CAGTTGAAAT AACCGCGTCT TGAATATAAG	12600
GAGAAATATT ATCCAGTTCC ACAAAGTCC CTTTTTGCCC GAATAACTTT CCGACTTCTT	12660
TTCCTGATTC ATCATAAAGC CGTGTTGATT CACTTAGCCC TGATTTTAGG GTCTCAACGT	12720
TTGCGGACTT GGCTAAGTAA AATAAATAAA TACTGGCACC TAAAGCGACT ACCAAACCAA	12780
TTAACAAGAA AATTTTATTG ATTTGATACT TTTTCCAAAT CCGCTTTCTG GCTTGATGGA	12840
ATTGTATTAA ATAAGGCTTT	12860

## (2) INFORMATION FOR SEQ ID NO: 145:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9086 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

AACGCTAAAA CCTTTAATAA ATAAAGGTTT TAGCnTTTTA ATAAAGAGCC ACCTGCCGGG	60
CTTGAACCGG CGACCCCCAC CTTACCATGG TGGTGCTCTA CCAACTGAGC TAAGGCGGCA	120
CTTTCAAAGG CTAAACTAAC TCGAGGATTG CAACGAATGA AACCCATAGA GTTCATTTTC	180
AAAGGCTACA CTGCGACGAG TAATCGTCAA AGTTCCTTTT GAAAAATCGT ACAAATCTCG	240
TGCACAAGAA ATAGAATAAC GTATAGACGA AAGTTTGTC ATTGCTCTTT CGTAAAAAAA	300

GAAAAATGAT TTTTCTCTAA GAAAACTGT TGACAATCTC TTTTCTCTCT AGTAAAGTGT	360
AAATCAACAT CACAGATAAC TATTCATTCG TTGAAGAGAA GAGTAAATTT TGATGACCTT	420
CAAAAGAGAG TCCTGATTGG TGAmAAGGGA CAAGGTTTCAG AAAAGTTGAm GATGGCCTCT	480
GAGAATAAAT TGCTGATAAG TAAGCAATTT CGGCTAAGCA CCCGTTACCA TGTGCACCAA	540
GTTAATAAAG ACTTGATAA GAGATGAGAT ATAAAGAATC AGTTATATTT CAAAAAAGG	600
TGGTACCGCG AATCATTCGC CCTTTACTAG CAAAACGCTA GTGAAGGGCT TTTTGCTGTG	660
AATCACTACG ACTGGCGCGA AGACAGAGTA GATGATGAAC AGTACAAGAT GACCGAAGGC	720
AAGTTGTGTC ATTTGCACGA ATCACTACGA CTGGCGCGAA gACAGAGTAG ATGATGAACA	780
GTACAAAATG ACCGAAGGCA GGTGTATCA TTTGCACGAA TCACTACGAC TGGCGCGAAG	840
ACAGAGCAGA TGATGTCTGC GATTACTCGA CGCAGGATGG CCTTTGTGAG TGATGAACAG	900
TACAAGGTGA CCGAAGGCAA GTTGTGTCAT TTGCACGAAA GTGTTCCTAC GATTCTCGTC	960
ACAGTCAAAT CCTGTTGAGC ACTACGACTG ATGCAAAGAG AGAGAGTAGC TATTATCTGC	1020
GATTACTCGA TGCTTGAAGG GGTAGTAACT CTAGTCTGTT TGATTTTGAA AATAAACGAA	1080
AATCAGTTTA AAATGAATAG AAATAAAAA GCAGTGATAT AGTAAGTACA AGACTTGATT	1140
TGTCTAACAG AGAGCCTAGG TAGCTGAGAA TAGGTGACAA AGACGTTTTG GAAAATGGCT	1200
ATTGAGCAAG CGCTCCGAGT GTTAAACACA AGGCGGCGAC GGGGGCTCCC GTTaTCGAGC	1260
TAGGGTATCG AACATATTGT TCCGTACTTG AAAAGGGAAT TTGTGTGAGC AAATTCTGAA	1320
CATGAGGTGG TAACACGCAA AAAAGsGTsC TCAACGATAC AACTCGAGTT GTGTCGTTGA	1380
GGATGCTTTT TTCTTTTTTA TGAAAAAATG CAGGTTGGTA AGAATGAAGG GGGATAAAGA	1440
ATGATTGAAC TAAAAATAT CAGTGTGACA TTTCAACAAA AAAACAAGA AATTCAGGCT	1500
GTCCAAGATG TTTCTTTGAC AATTGATAAA GGGGATATCT ACGGAATCGT TGGTTATTCC	1560
GGTGCTGGAA AAAGTACGCT AGTACGGGTA ATTAATTTAT TGCAACGACC AACAGCGGGC	1620
ACCGTTATTA TTAACAAAGA AAATATTTTA ACTTTTCTA AAAAGGAATT ACGACAACAA	1680
CGAAAAAAGA TTGGAATGAT TTTTCAACAT TTAAATTTGA TGAAAGAGCG GACTATATTT	1740
AGTAATATCG ATTTTTCATT AAAATATTCT GGGCTAAGTA AAAGTGAACG GCGTCAAAAA	1800
ATCAGCCATT TATTAGAGCT AGTTGGACTA AGTGAAAAAC GGGACGCTTA CCCCAGTCAA	1860
CTATCAGGAG GGCAAAAACA GCGGGTTGCG ATTGCCC GCG CTTTAGCAAA TGATCCTGAA	1920
ATTTTACTTT GTGATGAAGC GACGAGTGCG TTAGATCCAA AAACAACGGG ACAGATTTTA	1980
GCCTTGTTAA AAAAGTTAA TCAAGAACTG AACTTGACGA TTGTCTTAAT CACCCATGAA	2040
ATGCAAGTGG TTAAAGAAAT TTGTAATAAA GTCGCGGTGA TGGAAAATGG CTGCGTTGTC	2100
GAATCAAACG ATATTGTTTC GATTTTGTAGT CAACCCCAAC AACCTTTAAC CAAAGATTTT	2160
ATCCGGACAG CAACTCATAT TGACCAAGCA TTAACAACAA TTCTTGAACA TCCTAAATTA	2220
GCTGATTTAG ATAAAAATCA AGAATTAATT GAATTTTCTT ATGTGGGAGA TCAAACGAAT	2280



GAGCCATTAA TTGCACAATT GTATAGCCAA TATCAGGTGT ATACGAATAT CCTTTACGGC	2340
AATGTCGAAA TTGTTCAAAA TGTTCTATT GGCCATTAA TTGTGGTGCT TTCTGGTGAT	2400
GAGGCCCAGC GTCAACAAGC CCTTACTTAT TTAGCGAAAC AAGGCGTGC G AACCAATGTT	2460
TTGAAAACCT ATCAGCAAAC GAAGCAAAAG CAGAATTAC AGGTGATTTA AAAAAGGAGT	2520
GAAGTGAAG TGTATCAATT ATTTGAGAAA TATTTTCCGA ATGTTGTCCA ATTAACAA	2580
GAGTTTCTTC AAAGTACATG GGAAACATTA TACATGGTTT TTTGGACAGC ATTGATTGCT	2640
GGCGTGTTAG GAGTCTTGTT GGGTGTCGTG CTTGTTAGTA CTGGCCCCAG TGGTGTTTTG	2700
AAAAATCCAC CCCTGTACAG TGTCTTAGAA AAAATTATTA ATGTTTGCCG CTCTATTCT	2760
TTCATTATTA TGCTCGCACT GATTCAACCA TTAACACGAA TTTTGGCAGG AACGACAATT	2820
GGTACAACCG CAGCGTTAGT CCCATTAGTT ATTGGCGTAA TCCCGTTCTT CGCGCGCCAA	2880
ATTGAAAATG CGTTATTAGA AGTGGATCCT GGCGTTATTG AAGCGGCAGA AGCCATGGGG	2940
ACGAGTCCCT TAGGGATTAT TTTTAGGGTT TATCTAATTG AAGGGTTACC AAGTATTATT	3000
CGTGTTCAG CGGTGACAAT TATTAATTTG ATTGGATTAA CAGCCATGGC AGGAGCGATT	3060
GGAGCCGGTG GTCTGGGCAA CTTAGCGATT ACTCGAGGAT ACAATCGGTT TCAAACCGAT	3120
GTGACATTTA TGGCCACGTT AATTATTTTA ATTATGGTAT TTATCAGTCA AGCCATTAGT	3180
AATCAATTAA TCAAAAAAAC ATCAGATTAG AAAAAGGAG AAATGACAAA TGAAAAAATT	3240
TAGTAAATTA ATTGGACTTA TTGGGGTATT AGCTTTTACG ATTGCAGGTT GTGCATCGGG	3300
GTCTGTGAAG GATACTAAGA CAGAAACCGT TAACTAGGG GTTGTAGGAA CAAAAATGA	3360
TGAATGGGAA TCGGTCAAAG ACCGTTTGAA AAAGAAAAAT ATTGATTAC AATTGGTAGA	3420
ATTTACAGAC TATACGCAAC CAAACGCAGC ATTAGCAGAA AAAGAAATTG ATTTAAATGC	3480
CTTTCAGCAT CAAATCTTTT TAGACAATTA CAATAAAGAG CATGGAACGA AATTAGTATC	3540
aATTGGCaAT ACAGTCAATG CaCCATTGGG AATTACGCT AATAAATTGA AAGATATCAC	3600
GAAAATTAAA GACGGCGGAG AAATTGCTAT TCCTAATGAC CCAACGAATG GCGGGCGGGC	3660
GTTAATTTTA TTACAACTG CAGGACTGAT AAAAGTAGAT CCTGCGAAAC AGCAACTACC	3720
GAAGTGTAGT GATATTACTG AAAATAAAG CCAATTGAAA ATAAGTGAAT TAGATGcTAC	3780
GCAACAGCG CGCGCTTTAC AAGATGTCGA TGCTTCAGTG ATTAATAGCG GCATGGCTGT	3840
CGATGCTGGG TATACACCAG ATAAAGATGC TATTTTCTTA GAACCTGTAA ACGAAAAAGC	3900
GAAACCTTAT GTGAACATTG TCGTGGCCCG AGAAGAAGAT CAAGAGAATA AACTTTATCA	3960
AAAAGTTGTA GAAGAAATC AACAGAAGA AACGAAAAAG GTCATTGCAG AAACATCAAA	4020
AGGCGCCAAT GTTCCAGCCT GGGAAACATT TGGTAAAAA TAAAGGAGGC ATTTATAATG	4080
AGTACAACAA CGATTCAAAC AATCCAAGAA GCTATTGCTA CAGAAAAGGA ATGGATAATC	4140
CATTTAAGAC GTCATTTTCA TCAATATCCT GAAGCAAGTT TAAAAGAATA TGAAACGATT	4200
AAGCGAATTA AAGAAGAACT ACTAGCCTTA GCTATTCCTT TTGTAGAAGT AGGGGAAACG	4260

GGTGTTTTAG CAACCATTGA AGGAGGTCTT GGCGCTGGCA AAACGATTTT GTTACGTGCA	4320
GATATTGATG CGTTGGAATT GCCAGATGCA ACAGGTGCTG CCTATGCTTC TAAAAATCCA	4380
GGACTCAATC ATGCTTGTGG ACACGATGGT CATGCGGCAG CATTGCTAGG TGCAGCTAAA	4440
GTGCTCAAAA AACATCAGGA TACCTTTTCA GGAACGATTA AACTGGCCTT CCAGCCAGCT	4500
GAAGAAATTG GTGCAGGCGC TCGACAATTT GTGGAAGGAA ATTATTTAGA AGCAATCGAC	4560
CAAGTGTTTG GgATTcATT AGATTCCAGT GTGCCGGTCG GAAAATTAGT CGCTACCAAA	4620
GGCGCCACCA ACGCCTCTTG TGATATTTTT AAAAtTGAAG TCAGTGGTCA AAGTAGTCAC	4680
GTGCCCCAAC CACAGAATGG CCGGGATGcy stTTTAGCAG CGGCCaGTAT CGTTGTGGAA	4740
TTACAAAAAA TTGTAGCTCG CGAGATTGAT CCTTTAGATT CTGTCGTAGT AGGAATTGGC	4800
GTTTTACAAG CAGGAACACG CTATAATATT GTAGCAAACC AGGCAACCAT TGAAGTACT	4860
GTTCGAACAT TTAGTCAGGA AACGCGCCAA TTTGTTTTAC AACGAGTCGA AGAAATTGCC	4920
CATGAAATTG CCCAGTCTAA TCGCACAGAA ATTGCTGATT TTTCTGTTTA TCGGCAGCG	4980
AATCCATTAA TTAATGAAGA ACAAGCAACC AACCGTGCAC AGCAGGTAGC CAGTGAAATT	5040
GTTGGTTTTG AAAATGTTGT GACCGATCAT CCTAAAAGTT TAGGGGCGGA TGATTTTGCT	5100
GATTACTTAG CGGTAATTCC TGGTATCTAT GGACGGGTCG GTTCACGAAA CCCTGAAAA	5160
CCTGCTACTC ATTTTGGACA CCATCATGAA CAATTTGATA TTGATGAACG TGCCTTACTT	5220
TTAGCAGCAG AGTACCACGT ACGCTATGCA TTAAATTATT TATCAGAATA ACAGGGATAG	5280
GGAGCCGACA AAATGATGT TTGTGGGCTC CTTTCGTTAT AAACCAACGT GGCTTGCTTT	5340
CTTGTCATAC TATGGTAGAA TACAAGCATA AGAAAGGTGA TTCACATGTT TTTTGTTTTA	5400
ACTGATAAAT TAACAAGAAA AGTAGCTCCC GATAGTTTAA GAACAGATGA AGCACTGTTG	5460
TGGAGCCGAG CAATTAATTT CTTCATCGCT TGATAGGACA TTTTTGACT AGAAAAGCTA	5520
GTCTTATTTG AGTGACGTA CAAGCGGTTG TCGGGATTTC TGTCAGGATG GACGACGTGC	5580
AGACGTGGTT CTTTTTTTGT TGTCCCAAAA TAATCAGTTA ACAAGAAAGA AGTGACGAAA	5640
ATGAGTATTT TAACTATTGA ACATTTAACG CATCGATTTG GCGAAAAGGT CTTGTATGAA	5700
GAGGCTTCAT TGCAAGTGAA TAAAGGGGAT CATTAGGCT TAACTGGCCA AAATGGGGTC	5760
GGCAAATCCA CCTTAATTAA AATTTTAACG GGGGAAGTAT TGCCaGACGA AGGAACGATT	5820
CAGTGGCAAA AAAATTGCAA GATTGGGTAC TTGGATCAGC ACGTTTCTGT AGAACAATCA	5880
CTAACGATGG TTGATTTTTT GAAACAAGCC TTTCAAGAAC TTTTGTATAA AGAAGCGAAA	5940
CTGACAAAGC TTTATGAAGA ATACAGTCAA ACGGCTTCGG AAAAActTTT AGAACAAGCA	6000
GGCAAGTTAC AAACAGATTT AGATGAAAGT AATTTTTACC AAATCGACAC GATTATTCAG	6060
GATTTAGCCA ATGGGTTAGG ACTACAAGCA ATTGGTTTGG ATAAAAAGTT AGGGGAGCTA	6120
AGCGGTGGTC AACGTTCAAA AGTGATTTTA GCAAAATTAT TGTTAGAGGC CCCTGATGTG	6180
TTACTTTTAG ATGAACCCAC CAACTATTTA GATGATACAC ATATTCAATG GCTTGTTTCG	6240

TATCTGAATA ATTTTGAAGG AAGCTTTCTA TTAGTCTCCC ATGATTATCA ATTTTAAAT	6300
GAAGTGACGA ATTGCATCGC AGACATTGAG TTTGGCAAGT TAACCAAATA TACTGGAAAT	6360
GTAGAAAAAT CTTTGCACA AAAAGAGCAG AATAAACAAA CCTATTGAA ACAGTATCAG	6420
GCCCAACAAG AAAAAATTGA AAAAATGGAA GCCTATATTC GTAAATACAA AGCTGGAAAT	6480
CGAGCAACGA TGGCTAAAAG TCGACAAAAA CAATTGGACC GGTGGAACG ATTGACTCCG	6540
CCTGGTTCCT TGACTAAGCC AGCGATTGAA TTTCTTATC AAGGGTTAGT TGCAACGCAA	6600
GCACTAACGA CCCAGAAGTT AGTTGTCGGC TATCGGGAAC CTTTGTTAGA ACCGTTAGAT	6660
TTAATGGTTC ATGTCGGTGA AAAAGTCGCA TTGAAAGGCT TTAATGGGAT TGGCAAATCA	6720
ACATTAATTA AAACGTTGAC GAAAGTGATT CCTTCATTAG ATGGAGAATT TCATTATCCG	6780
CTGAATACAA AAATTGCTTA TTTTACCCAA GACTTAGCGT GGCCTAATGA GCAGTTAACG	6840
CCACTAGACT ATTTATCAGA TCGTTTCCA GATACAACGA TTAAAGAGCG AAGAAGTCAT	6900
TTGGCTAGAG CAGGTTTACC AGATAAGTTA GCAATGCAGT CGCTAGCCCT GTTAAGTGGT	6960
GGTGAGCAAA CAAAaGTAAA ACTAGCGGAA CTAATGATGC AAACAAGTAA TTTGTTATTT	7020
TTAGATGAGC CAACCAATCA TATTGATGAA GCGGCCAAAA AAAGTTTACA AGAAGCAATT	7080
CACGTTTACC CAGGAACGGT TTTTCTGGTT TCCCATGAAG CAGATTTTGA TGAGGAAATT	7140
GTGGATCGAG TAATTGATAT TGAGGAATTA GTTAAATAAA GAAGAAAGAA CCAAGTTGAG	7200
GGGGAGTCAA CTTGGTTCTT TCTTCTTAT TATTTATTIA CTCATTAAGC CATCATGGAT	7260
TTTTGTAAA TTCCAGTTCA TCATGCTGTA GTACGTATCG CCTTCTGTTT CTTCTTTGGC	7320
AAGAGAGTCT GTGAAAAGTG TATCGTAAAT TGGTCGTTT ACTTCTTTTG AGACCCGTTT	7380
CATACTACGT TTATCGACAC TGGTTTCAAC AAATAACACA GGTGCTTTTG ATTTCTTAAT	7440
GGTATCAATA ATCGTGGTCA TTTGTTCAAG TGTyCCTTGA CTTTCTGTGT TAATTTCCCA	7500
AATATAAGCG GCATTTAAAT CATAAGCTTT GGAGAAATAT TTAAAGGCAC CTTCACTTGT	7560
AACTAATAAT TTTTATCAT CAGGAATATC AGCAAATTTA GCTTTGGCTT CCTCATGTAG	7620
TTTGCTAAGT TTTTCGGTAT AATTTTTCGC GTTTTCTGTA TAGAAATCTT TATTTTTTGG	7680
ATCTTTTTCT ACTAACACGT CACGAATGTT TTCTACATAT TTAATGCCaT TTTCAATGTC	7740
TAACCAAGCA TGTGGATCTT CTGTTTGTTT tTGACCGGCA CTTGTAAAT ATTGTGGCGT	7800
AACATTTTTG CTTGTAGAAA AGTAATCTTT ATTCTCAACT TTTTGGCCG TTTTCATTAA	7860
TTTGTTAAAC CAGCCATTTC CGCCTGTTTC TAAGTTCAAG CCGTTAAAGA ATAAAATGTC	7920
CGCTTCAGAA GCTTTCGCAA TGTCTTCTGG TAACGGTTCA TATTCGTGAG GGTCTGTCCC	7980
AATTGGCACA ATACTATGCA GCTCAATTTT GTCTTGCCCA ACATTTTTCA CTAAATCAGA	8040
TAGGATCGAG TTCGTTGTCA CAATTGCTAA TTTTCTTTC TTTTCAGCGG CTTGATTCCC	8100
GCAAGCAGCT AACGTTAACC CTGCTAAAAG TGTTAAAAA AATAAACTAA ATTTTTTCAT	8160
TCGTTGATTC CTCCATTCTT TTCTCACGGT TTACAAATAC TAGGCCTTTC TTTGGTGAGA	8220

AAAAGAAAGC CAATAAAAAG AATAAGGCGG CGGTAAAAAC AATCGTCGCA CCAGAAGCTA	8280
AGTTATAAGA ATAACTGAAG AATAGGCCAA TCACTGAAGT TAAAAATCCA AaaGTaGAGG	8340
CTAaGCCAaT CATtGTCGGT AAGTGATTTG TTAaTwGATA gCTGTTGCTG CTGGCGTAAT	8400
CAGCtGGCAT GACTAAAATA GTCCGACGG TTTGTAAmGA AGAAACAGCG ACTAGTGTTA	8460
GTAAAaACAT CAATGCATAG TGGAAAAATT GAaTATTTAA ACCGTAGGCC TGAGCCATTG	8520
TGGGATCAAA AGAAGTGAGT TGCAACTCTT TATAAAACAA TGCCACAAAA ATCAAAACAA	8580
TCACACCGAC GACACAAGTA ATTAAATAT CGGTATCTGC GACCGCTAAT ACATTTCCAA	8640
ACAGAATATG ATATAATCG GTGGAACTTT TTGCAAAAGA GATAAAAATA ATCCCTAAAG	8700
CAAAAAAGA ACTAAAGACA ACACCAATGG CAGTATCATT TTTTAGTGGA CTTTTTTGGG	8760
TAATAAAGCC GATCGATAAT GCAGCTAATA AGCCAAAAAT AGAAGCGCCA AAAATATAAT	8820
TGAAACCAAA CATmTAAGAA ACAGCGACCC CTGGTAAAAAC AGCATGTGAA ATCGCATCAC	8880
CCATCAAAGA CATCCCACGT AAAATAATGA ATGAACCAAT TACCCCTGAA ATGAGTCCGA	8940
CAATTATAGA CGTCAAAGG GCATTTTGTA AAAATTGATA ACGGAATAAA CCATCAATAA	9000
ATGCAGCAAT CATTTTGCAA CCCCTGTAT TAATAATCA CCTAACAAAT CACCGTATGC	9060
TTCTTGAGC GTTCTGCAG TAAATG	9086

## (2) INFORMATION FOR SEQ ID NO: 146:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8546 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

TGCGCTTGCA ATAGTGAAT GACTAATCG TTTATTTTCC ATGTTCATCT GATAATCAGC	60
CCCCTAACAT TTTGGACACG CTTTCATTAT TACATTAAAG GGGCGTTCCA AAATAAATCT	120
TTTATTTTTG GAACGtCCAT CTCTAAATCA TTCTCTTAA AACTGTAATA GCTTACGACG	180
ACAAGTCAGA ATTATTCATC CTCATCATCT ACTGCAACAT TTAGCAGCAA TTCGTTGACT	240
TTCACAATAC TTTCTGCTGA TTCCTTAACA TAGTCACCTT CGATCGCTAA TAAACTTGCA	300
TTTACATAGC TGATAACCAT CGGTAAATTC ATCCAGTAT ATAAGTCAAA CGAAGCACCT	360
TGCATAATTT TTTTGATAA AATATTACAA GGTGTCCCC CTAGTAGATC AGCAAAGACA	420
ACAAAATCTT CTAGCGATTC CGTAATCGCT TCAAATTTT CTAAAAAGTC AGCTTCACCT	480
tCTTCTGGTA ATAAAGCAAC AGTGGnAATT GTTCTTGCG GTCCATAAT CATCTCCGCA	540
CTTTTTTTCA ATTCTCGCA AAAACGGCCA TGACTTACTA AACTAAATG TTCACTCATC	600
GATTTTCyCC TGTCTTACTT ATTTTGAAwT AACACCTAAC GCAGACAaTG CGACACAGAC	660
AaTTAATACG ATAAAAATGG CTTTAGTAGA CGTCATATTC TTTTACCTA GCATCCAATA	720

AACGCCACCG	ACAATTCCTG	CAGGAAGTAG	ACGTGGTAAA	ATCATATCCA	AGTTATTTTG	780
CATGTTAAGT	GTTACATCCC	CAATACTTGG	TGCCCATGAT	AGTTTGACAT	TAACCATCGT	840
TGCCACTAAA	GCGCCAACCA	TGAAGACCCC	TAGTAAAGTT	GCAGCGTCTG	TTAAGGCAGT	900
TAAGCGGTGT	TGCATAGTTG	TTACCAAAGA	AATTCCTTCG	CGATACGCAA	ATTCCAATTG	960
TTTCCAACGG	AAGACCATCA	CGGCAATTG	AGCAACAATC	CAGATAAAGA	TACCAGTTGG	1020
GTTACCATTA	ATCGCCATGT	TGGCaGCTAG	CGCCCCAAAA	ATAGTTGGAA	TTAAAGCTGC	1080
GAAAATTGAA	TCCCCAATG	cTGcAAACGG	TCCCATTAAA	CCAGTTTTCa	AACCTGaAAC	1140
TGTTTgTTTT	CCAGCAATGc	CTTCTTTTTc	TTCaATGGcC	AAATCGATCC	CTGTAAtAAT	1200
TGTATTAAAG	AAGTTTGAAG	TATTGAAAAA	TTGCGTGTGT	GTTTTcATTA	CTTCTTTTAA	1260
TTCAGGCGTA	TCATCCCCAT	AAATTTTACG	CAATTGCGGC	AAAATCGTAT	ACAAGTAACC	1320
TGATCCTTGC	ATTCTTTCGT	AGTTCCAACC	TAATTGGAAA	GTAAACAAGC	TTCTGCGATT	1380
AATTTGTTTA	AAATCTTCTT	TCGTCAATTT	ATAATTAGAT	TTCGTCATCT	TCAATCTCTC	1440
CTTCGTCTGA	ATCATTCGCT	GCATGTTCTT	TTTTGGCTGC	TGGCCCACTC	GGAATCATTT	1500
GACCATTTTT	GTAGCTAATT	GCGGCTAAAG	CGAAACCAAT	TAAAGCGACT	GCTAACATTG	1560
GTAGTGCGTT	AAATACACCA	CTGAAGTCTT	TCACAACGCT	CGCAACAGAT	GTCCTAAAA	1620
GTTGCATGTT	TGTAAAGATT	GTTCTAGTA	AAGCAGTTAC	AGTAAAACCA	AGAATTAAAT	1680
AAGGGAAATG	TTTTTTCACT	GGTAAGTAAC	GTAGTAAAT	TGCAAAACCG	ACCGCTGGTA	1740
AGACCGCCCC	AGCAACAGAC	AAGCCATCAC	CTAACCATTT	CAAATCGCCA	TTTAGGACAG	1800
ATACAACTTT	TTGTACTAAA	CCACCACCAA	AAGCAAGTGC	TAAGAAGACT	GGAATCATTC	1860
GAGAAAGAGA	CCAAGGTAAg	GCACCCATTA	AGAAGTTACG	TTCAATTCCT	TTGTAGTTCA	1920
TATCTTCGAC	CATTTTATCA	ATCCGATGTG	CAAAGTAGGT	ATTCGCAAAA	CGAGCCAAAA	1980
TGTCTAATTG	AATCATTAATA	CTAGCAACTG	GTA CTGCAAT	AGCAGCGATT	GCTTGTTcAG	2040
GGTTCATTCC	TAAAGACACA	GAAAATGCTG	TGGCAAGAAT	CGTGCCTGAG	TTAGCATCAA	2100
TTTTGGAGGC	ACCCCCGAAA	GTCCCAACTC	CTAATACAGT	TAATTGCATG	CTTCCACCAA	2160
TAATAAGTCC	TGCTTTAATA	TCCCCCATCA	CTAATCCTGC	AAAAAGACCT	GCAAACACTG	2220
GCGCACTTAA	TGAAGAATAA	ATTTGTAATT	CATCTAAAAT	TTGGTATCCT	GCGTATAATG	2280
TTAATAATAA	GATCTGCCAC	CATAAGATAT	TCATTTTTTG	TTCCCTCCTA	TTTTGTTATC	2340
AAACTCATAA	AGTCTTCCGC	TTGTGCACTT	GGCACCATTT	GAGCAATCAG	ATGGACACCT	2400
TTAGCATTTA	ACTCTTTAAA	AACCTCAATA	TCTTGATCGA	CCACATTGAT	TGACTTAGTA	2460
ATCGACCGTG	TTTCGTTTGT	TTGAGACATG	TTCCCAACAT	TGATTTCTTT	AATCGGCACC	2520
CCAAGTTCTA	CTAGTTTTTA	TAAACGATCT	GGTTTCCGTA	CCACAATTAA	TAAACGTTGT	2580
GAGTCGTATT	TTCCTGCTAA	AATATTTTGA	GCAGCTTTTT	CAACTGGTAA	CACGCTTAGT	2640
TTCACGCCAG	CTGGTGTTGC	TAATTTCAAG	CCACTTTTTT	CAATGGCATT	CCCTGCAACT	2700

TCATCATCAA	CGACCATAAT	TCTGGAAATA	TTCAATTTTG	TCGCCCATAA	ATTGCCACT	2760
TGTCCATGAA	TCAAACGTCC	ATCAATTCTT	ACGCCAATAA	TACTCATTGT	AATCCCTCTT	2820
TCATTTCTTT	ATATAGTGGT	TCTTTTACTA	ACTGAATTTT	TGGTTGGTAC	GTTCTTCCG	2880
TTTCAAAAAT	GACGATCTCA	TTTTGCCCTT	CTTTTAATAA	TCCTTTGGGA	ATATATAAGG	2940
AAAGCGTCGG	ACCAACATTC	CAGAAACGAC	CAAGATTGGT	TTGGTTAACA	AAAACAATTC	3000
CTTTGCCAAA	TTTAGAAACA	TCAATAAATG	TATCTTTCAC	TTCTGCTAAC	TCCACATGAT	3060
ACTGGTAAAA	GCTTGGTTGA	TCAGGCTGCC	ATTCTCTTGA	ATAGTCTACC	TGTTACAGC	3120
TTGTCATAGG	TAAACAATAT	TGTTGCCATT	GCGTCATAAA	ATGTAAATCT	GCCATGACAC	3180
CCGTTCGAAT	CCCTTTTTTC	TGAGTATCTG	CAAATAGTTT	ATGCCCGTAA	TTGACACGAC	3240
CCATATTTTC	CATTAAATA	TCGATTTGAT	TATTCTCTTG	CGGCAGTGTA	ACGTAAATAT	3300
CTTCGCCAAT	TTCCGTTTGA	TATTGCGTTG	CTTGATGTAC	CTGATTGACA	AATAATTGCA	3360
GACGATCAG	ACCGTCAATC	ACTCTTAGCT	TTTCTTCTGC	AGCATCTTTT	TCAATGCTTG	3420
tTCGGTAAAG	AAGATAGCCg	GTATTTTGTC	CTAATTGTTT	CATGGTTTGC	GGATAAACGC	3480
TGACCACTGG	CTGACTAATC	GTTTCTAAAG	TCGCAAACAA	ACTCACTTTG	TTGGTTAAAG	3540
GAATTGCTGT	TTGCGCAAAA	GATTCTTTTA	CTAATGGTTC	AGCTTGTTGG	AACGCAGGGT	3600
ACTCTTCGTG	AAGCATTTTT	TGTAACGCAA	AATACTTTTC	GGTTGGATTG	CCTTGTTTCT	3660
CAAGTGGCGC	ATCATAATCA	TAGGAAGTGA	TTTGTGGTAA	ATCAATGGTT	CCTCGTGCGG	3720
AGCAGCCATT	CATAAACCCA	AAATTGGTGC	CGCCATGAAA	CATATATAAG	TTTATACTGC	3780
CTAATGCTAA	TGCTTCTCTA	ACTGATTCTG	CCAATTCTTG	CGGATCTCGT	TTAATAATTG	3840
GCTCTTTCCA	ACGATTAAAC	CAGCCATCCC	AAAATTCCAT	ACACATTAGC	GGCCATTTTT	3900
TACCATGTTT	TTCAAAAAAG	GCTTGCATCA	TACCGAAATT	TTCTTTGGCT	TTTGAGCCAA	3960
AATTTCCGGT	TACTAAAATA	TCATCTTCAA	TCATGCTGCC	TGCCCGCAAT	GTTGCGCGCC	4020
ATGGCCCGTC	AGAAGTAAAA	AATGGTGCTG	TCACGCCGCG	TGCAATCATC	AAATCGCGAA	4080
TCGCCCGTAA	ATACGCTTTT	TCTTCGCCAA	ATGAACCATA	TTCGTTTTCG	ATTTGAATCA	4140
TTAAATATT	GCCACCGTTA	GCAAGTTGAT	GCGGAACAAT	TTTTTCCATT	AAAACATCAT	4200
AATATTCCGC	AACATGCTTC	AGATAGGTTG	GATTATTCTG	ACGCATCCGT	CCTGGTTTCT	4260
TTAATAGCCA	GGCTGGAAAA	CCGCCAAACT	CCCATTCTGC	ACAAATATAA	GGAGACGGCC	4320
GCACAATCGC	ATATAGTCCA	AGTTCTTGTC	CTAATTTTAA	AAAGCGCTCC	AAATCTAATA	4380
TTCCCTCAAA	ATGAAATGTT	CCTTTTTGCG	GTTTATGTAA	ATTCCATGGT	ACATATGTCT	4440
CAACTGTATT	GAAGCTTAAA	GCTTTCAAGT	TATAAAGGGA	ATGATGCCAA	TCTGAGGGAT	4500
CTACACGAAA	ATAATGAATG	GCGCCTGATA	AAATTTTAAA	TGATTGACCA	TTTAAAAGAA	4560
ACTCTCTTT	AATTTCAAAT	CTGTCCACAT	TGTCACCCTT	TCTAGTAAAC	GCTTTCAATT	4620
GCTGGATAAT	TGCTAAATAA	AGAAAGCTCT	TTATTTAGCA	ATTCGACTA	CTTGGATGTT	4680

CATAAGATGT	TTCTTCACGT	TCTCTACGTC	TACATGCCCA	GTCATTCTTT	CTTGGGCATT	4740
AGCCATTCCA	GCCGCCATGC	CCCATTTTAG	CAGTTCGGCA	GCTGGGGCAT	CCTTGGCTAG	4800
TCCATATGCA	AGACCAGCAA	TGGTTGCATC	TCCTGAACCA	ACAGGATTTT	TwGcTtGAaT	4860
CGTTGGAATT	TTTACCACGA	TAGAATTGAT	CGTGATGTTT	GGCAATTGCA	CCATCTTTTC	4920
CTAAAGATAT	GACAATCCAT	TCAATTCCCG	CAAACATTGG	TTTTGTAA	GCTGTTTGT	4980
CAGCTGCTAA	TGGATTTTCT	GAAAAATCTT	GCCCTAGCAA	TCCTTCTAGT	TCCTCTAAAT	5040
TAGGTTTAAT	TAAATACGGT	TTCCATGGTC	CTTGAAGAAC	TTGCCTTAA	CTATCACCAG	5100
AAGTATCTAG	TAGCACTTTA	ACCTCTTGTC	CGTGAnCTT	TTGAACATA	TCTTGATAAA	5160
AATCTGATGG	TAATCCTTTC	GCTAAACTGC	CAGAAATTGT	GACAATTTCT	GCTTGCTTAA	5220
TTAATTGATC	GAAATTTTCA	AGGAAGTTAC	TGATTTCTTC	TGGGGAAACA	GTTGGTCCAG	5280
CTTCTAAAAT	TTCAGTTTGA	TTCCCCTCGT	GCAAATCGC	AATCGAATCA	CGTGTTCCT	5340
CTTTGATTGA	CGTAAATGCT	TGGGGGATGT	TTGCTTTCTT	TAACTCATTT	GCAATAAACG	5400
CCCCATGAAA	ACCACCAAGA	ACGCCGGTAG	CGATAACATC	TCCACCTAAA	TCATGAATAA	5460
CGCGAGTGAC	ATTTAGTCCT	TTGCCACCAG	GGGTTTTGGT	CACCTGACTG	GTCCGATTAA	5520
CGGTATCAAG	CTTTAGATGA	TCTAATAGGT	AGGAGATGTC	AATTGACGGA	TTCATTGTTA	5580
CTGTTACTAT	CACAGCTATG	CCTCCCAATT	ACCCTTCGTT	TACTTGAAC	TTCGCATGCC	5640
AAGAGCTGGC	GGTTGCTGCG	ATGACTTCAT	TTAAACTTTC	AATGTTTTCT	CGGCCTTCTG	5700
TTTTTAACCA	ATCACAAGCC	GCAGTTTCTC	CTGCTTCAAC	GAAAGGTTTC	ACGCCATTTT	5760
TCCACGTAGC	ACGCCACAT	AAAACACCAT	TAAATGTTGA	ACCAGCTTCT	TTAGCGAAAC	5820
GTAGTGTTTC	TTGGAATAGC	TCTGTTGAGA	CACCGGCACT	TAAAAAGATA	AATGGTAAAT	5880
CAGTGGCTTG	ACTTTGTTCT	AAGAAATAAT	TCGCCGCTTC	CTCTTTTGTA	TACGCTGTTT	5940
CAGCCGGCGC	AAAACCTTCT	ACAAAATTCA	TATTAACCTG	AACTTCCACT	TTCAATACAT	6000
CTACTTTATA	TTGGGGTTTA	GAAATTTCTT	TCATCATTTT	ATTCACCTTA	TGCGGTTTAA	6060
CTTTTGCATA	TTCTAAAGAA	GTCGCATCTG	CTATTTGTGC	ATCATAAGAA	ACTAATTCTA	6120
AGTAAATGG	TAGGTCTTCT	TCTGCACATT	CACTTCCTAA	GCGTTCAATG	AATACATGTT	6180
TTGATGATT	GATTCTGGA	TCCTCATCCA	CATCATAGTA	TAGTAAAAAT	TTAATCGCAT	6240
CTGCGCCTTC	TTCTTTCAAG	CGTAAACAG	ACCAATCTGC	TAGTAAATCA	GGCAATCGTC	6300
CCGGAGTTGT	TGCATCATAG	CCCGTTTTTT	CATAAGCGAG	TAACAATCCA	GCCTCGGTAT	6360
CGCGTGCTTT	TGCAGCAGGC	AATCCATATT	CTGGaTCCAA	TAAATGGcA	GACGCATAAG	6420
GTGTCAATTC	TTTCGAGACT	AATTCTTTAA	ACGTTTCAAT	TTGTGCATCA	GTTGGCTCTA	6480
CATCTAGCGC	TTTGATCATT	TTTTTCAAGG	CACCTCGTTG	ATCAATGGCT	AATGCGCTAA	6540
TGATTCCTTC	TTGGGTTGAC	AAACGATCCA	TTGCCGCTTT	TTTTCCAGCT	GTTAACGTAA	6600
GCATAGTTCC	TCCTCCAAAT	GTTGTATTAT	TTATAGTCAT	AAATCGTTAC	GCCTTTAACC	6660

ACACGATTCA CTGTACCTGT CGGTGAAGGG GTATCCGGTT TATTTCCAAC TTTAATAGAT	6720
GTTAAAAGTG CGACTGTTTG GGCTACCATG ATATCTGCnA ATGCnAAATA GCCATCAGGT	6780
AATAGAGCAG TTTCTGCGGa ATATTCAATm GyTGAACcTG AGAAATTACG TTTGCCAGGT	6840
TGAGCAATCG CAAAAACACC CGCCGCAATT TCGTCACCGT GAACTTCTTC TAAGATATCC	6900
AAGTCATAAT CTCTTGTTGA TGGTGTATTA TTaACAAAAC CAAACATTAT TGTTTTTTCA	6960
TTAATAAATG ATTTTGGTCC ATGACGGAAG CCCATAGAAG AGTCAAAAAT TGTGGCAACT	7020
TTGCCGGCTG TTAGTTCTAA AATTTTCAAT TGCCTTCTC TTGTCAAACC AGCTAAACTT	7080
CCAGAACCAA CATAACAAT GCGTTCAAAA TCAAGGTTAA CAATTTCAAT TAATCTTCT	7140
TCACGGGATA GCACTTCTTC ACCTAATTTA ATTAGAGCGG TTACATAATC TTGTTTTTGC	7200
TCAAAGGCTG TTTGATCAAA GAGTAGTAAT GTACCCAAAG TCATGCAAGA AAAGCTACCA	7260
GTCATCGCAA AACCACCATC ATTGGAACGT GTTGGCATTa AGAATAAAAA GCTGTTTTCT	7320
TCTTTTTGAC TTATTTGTGC TAGTTGTCCT TCTTCGGCAC AAGTAATCGC TAAATGATGA	7380
ATTGTTTCGA CCACTTGATT AGCTACTTCA ACTGCCGCTA AACTTTCAGG ACTATTCCa	7440
CTTCTTGCAA ACGAAACTAA CAATGTTGGT TCGTCTTTAA TCAAATAATC TTCTGGTTTG	7500
GCCACAATAT CCGTTGTCCC AATACTTTTCG AATGAAAAGG CTTGTGTGTC GCCGTGTGCT	7560
CGGAGATAAG GTACAACCGT ATCACCGACA TATTGCGAAG TCCCTGCTCC TGTA AAAATT	7620
ACTCTTGTC GTTTTCCTTG CGCTTTGGCC TGGACTTCTT TTA AAAAGTT TTCCAAAGCT	7680
GTTTGATTTT CATGATAAAG TGTCACGGTT TCTTGCCAAA GCTCAGGTTG CTGACGAATT	7740
TCACGTGTAG TGATTTCCGC TCCTAGTTGT TCTAGTTCTT CTTTTTCTGC TGTAACATA	7800
CTTAActCTC CTAActATTC CGAATATGGC GTATCTGATA ATGAAATTGa TCTGCTCGCG	7860
CCACACTCAA CGTAAATTCA ATAATTTTCA TTTTCATGTT ATATGTCTGG CGTGCTAAAT	7920
GAAGGACTGG CGCTCCTTCC GGaATCaTCA ATAATTTGGc ATCTTCTTTT GACGCAATGC	7980
TAGCGTATAA TTCTTCATCA GCTAAACGAA CGGTTTGCTT AAAATCTTCC GAAAATAGAT	8040
CATATAATGG TTTGCTACGT AGTAATTGAT CCGTGAGTGA TAAGAAAAAT TTCACAGGCA	8100
AGTATGTGTC TTCAACCATT AATGGCTCGT TATCAGCAAT CCGCAACCGA CTTAATTTAA	8160
AAACTGCCTC TCCTAACGAA AGATTCAAAT GCTGACAAAT AAATTTATCC GCTTCCAATT	8220
TTTCGAAGGA TAATATACGG GTATGTGGTT TTCGACCCAG CCCTTTCATT TGTCTGTAA	8280
AACTATATGC GCCAGCTAAA TCAGCGGCCT CTTTTTTTAT ATCTGAAACA AATGTTCCCTT	8340
TGCCATGACG ACGATAGATG GAACCTCTAT TTTCTAGTTC TTGTAGCGCT AACCGGACCG	8400
TCGTCCGACT CACTCCATAT TGTGCAGTCA ATTCACGTTc AGAGGGCAAC TTGTCATGAG	8460
GAATCATGCA TGTTTCGATT CGTTCCTGCA GCAAAATCAAC AAGTTGATGA TACAAAGCTT	8520
TGTTTTTAAA GGCATTTTCC ATAAct	8546



## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3493 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

CGACACATTG TGGGACATG CTGCACGTTT TGGACGACA GTTGATAATT TAATGTCACT	60
AAATGGATTA ACATTGGATT CTGTTTTATC GGTGGGTCAA ACCATCCAAA TCGGTTAATT	120
TTGCAATTAT TTAGTGAAG AAAGACCTWA CAAAGCAGT AGTTTTTGTA AGTCTTTTA	180
TCGTGAAAT AGGTAGGGCT AGCAGCAATT ATTTGATGAG CCTAATCCGT TTTTTATTT	240
AATTTTTCAC AAATAAAT TCGTTTTCT GGTAGAAATC GTTTTTTGA GCTATGCTTA	300
GAGAAAGAAT TATTATATTG TTTGTAAAT TGGAGGATGT AGGATTATGA CTAAAGGTTA	360
TGTAAATCA GTAACAGAGT TAATTGGTCA AAGCCAGTT GTGAAATTAA AACGGATGGT	420
ACCAGAAGGA GCTGCAGATG TTTTGTTAA ACTAGAGTTC TTAATCCTG GTGGCAGTGT	480
GAAAGATCGA ATTGCTTTAA GCATGATCCA ACAAGCCGAA GCAGATGGTC GCCTAAAACC	540
TGGACAAACG ATTATTGAGC CGACTAGTGG TAATACAGGC ATTGGGTTAG CAATGGTAGG	600
TGCTGCGTTG GGATATCCAG TTAAGATTGT GATGCCAGAT ACTATGAGTA TTGAACGCCG	660
CAAATTAATG CAAGCATAG GGGCTGATCT ATTATTAACC CCTGGTGCCG AAGGAATGAA	720
AGGGGCAATT GCAAAAGCTA CAGCATTAGC AGAAGAACAC GGGTACTTCA TGCCTTTGCA	780
ATTTAATAAT CCAGCTAATC CAATGGTACA TGAACAAAA ACAGGAAAAG AAATTGTTGA	840
TGTCTTGGT AACGTGGCT TAGATGCGTT TGTCTTGGT GTTGGCACTG GAGGAACCGT	900
TACAGGAGTT GGCCATGAAT TGAAACGGAT TTTCCAGAT ATTGAAATTG TTGCAGTAGA	960
ACCAACAGAG TCGCCTGTTT TAGAAGGTGG CGAACCAGGT CCACATAAAA TCCAAGGAAT	1020
AGGCGCTGGT TTTGTCCAG AAGTTTTAGA CACCACCGTT TATCAAAAAG TTGCCGCTGT	1080
TTCTAGTGAA GACGCATTAG AAATGCTCG TTTAATGGGG CCAAAAGAAG GTATCCTTGT	1140
GGGGATTTC AAGGGGCGC CAATTAAAGC TGCCATTGAT TTGGCAGTCG AATTAGGCGC	1200
AGGCAACCGT GTCTTAGCGC TGGTTCCGGA TAACGGTGAA CGTTATCTTT CGACAGCTCT	1260
TTACGAATTT CCAGAATPAA AAACACGCCA AAAACGAGC ATTTCTCTTA GTTTAGTAAA	1320
ATGCTCGTTT TTTCCACGAG AATCCTACAA TAATTCTTGT CAATCGTAAT AAGAAATAGT	1380
ATTATTTAAG AAAGATAGAC AGGGGAGGGA TACAAAGATG GACAACGTAT TGGTTAAAAA	1440
TGCACTTGCT GAATTAAAAG AAGCCAATAT CCGAATTACT CCGCAACGTT ATGCTATCTT	1500
GGAATATTTA ATCGAAAATC ATACACACCC AACAGCTGAT GAAATTTATC GCGCACTAGA	1560
AGATCATTTT CCAAATATGA GTGTAGCAAC GGTTTACAAC AATCTACGTT TATTTACTGa	1620
AATCGGTTTC GTTCAAGaAA TGAGTTATGG CGATGcATCT AGTCGTTTTG ATTTTAGTTC	1680

GAAGAAACAT TATCACGTGA TTTGTCAAAA ATGTGGTAAA ATCGTTGATT TTCATTATCC	1740
AGGGTTAGAG GACGTTGAAA TGGCCGCTAG TAAATTAACA GGCTTCGAAA TTAATGAACA	1800
TCGTTTAGAA TTATATGGAT TATGTCCAGA TTGCCAACAA GCACAACAGG AGAATGTGTA	1860
AAAAATTACA TACTGTAAAA AATTTATCAA AAGATAAGTC ACTTGTCTAG GCAAATTTCA	1920
CAAGCGTAAG AGTTTTCACT CTTATGCTTG TTTTTTTGTG TGGTTATTGA CTTTAAATT	1980
CTTAAAAACA AAAAATTAAC TAGCTATTAT TAAGAATTAG CAACAGTTTG TGAATTTTTT	2040
TATATGCGAT TCACAAGCAT AGAACTCGAA TAAGCTTGAT TCCTCTGATA TATTTTAAAC	2100
AGACAGAGGG AGACGTCACC TCTGCAAAAC AAAGATTGTG AAAGAAGGGA TTTACAACAT	2160
GAAAGTCGTA GTCGTAGGAT GTACCCATGC TGGAAGTTCA GCAGTTAAAA GTATTTTAGC	2220
AAATCACCCG GAGGCGGAAG TCACTGTATA TGAAAGAAAT GACAATATTT CCTTCCTATC	2280
TTGCGGAATT GCCTTGTAAG TTGGAGGCGT AGTTAAAAAT GCTGCTGATT TATTTTATTC	2340
AAATCCAGAA GAACTGGCTT CTTTAGGAGC AACAGTGAAG ATGGAACACA ATGTTGAAGA	2400
AATCAACGTG GACGACAAAA CTGTTACAGC GAAAACTTA CAAACGGGCG CAACTGAGAC	2460
CGTTTCTTAT GATAAATTAG TAATGACAAC TGGTTCTTGG CCAATTATTC CACCAATTCC	2520
AGGAATTGAT GCTGAAAATA TTTTATTATG TAAAACTAT TCACAAGCCA ACGTGaTCAT	2580
CGAAAAAGCG AArGATGCCA AACGTGTCGT TGTCGTTGGT GGCGGTTATA TTGGTATTGA	2640
ATTAGTAGAA GCTTTTGTAG AATCAGGTAA ACAAGTTACT TTAGTTGACG GGTAGACCG	2700
CATTTTAAAT AAATATTTAG ATAAGCCATT CACAGATGTT CTAGAAAAAG AATTAGTTGA	2760
TCGTGGCGTT AATTTAGCTT TAGGTGAAAA TGTCCAACAA TTCGTAGCAG ATGAGCAAGG	2820
AaAAGTAGCA AAAGTCATTA CACCAAGTCA AGAATTCGAA GCAgATATGG TGATTATGTG	2880
TGTTGGTTTC CGTCCAAATA CAGAATTATT AAAAGATAAA GTTGACATGT TACCAAATGG	2940
TGCAATCGAA GTCAATGAGT ATATGCAAAC AAGCAATCCT GATATTTTTG CTGCGGGTGA	3000
TAGTGAGTT GTTCACTACA ACCCAAGTCA AACAAAAAT TATATCCCAT TAGCAACGAA	3060
TGCTGTTCGT CAAGGTATGC TAGTAGGTCG TAATTTGACT GAACAAAAAT TGGCTTATCG	3120
TGGCACACAG GGAAGTTCTG GTTTATATTT ATTCGGTTGG AAAATTGGTT CAACAGGTGT	3180
GACAAAAGAA AGTGCCAAAT TAAACGGACT AGATGTCGAA GCGACTGTTT TCGAAGATAA	3240
CTATCGTCCA GAGTTTATGC CAACAACTGA AAAAGTTCTG ATGGAATTAG TCTATGAAAA	3300
AGGAACACAA CGCATTGTTG GTGGTCAATT GATGTCTAAA TATGATATCA CACAATCTGC	3360
CAATACTTtA TCATTAGCGG TCCAAAATAA AATGACgTCG AAgATTaGCG ATTTCTGATT	3420
TCTTCTTCCA ACCACATTTT GATCGCCCTg GGAcTACTTA AACTTATagc CCAAGCCGCT	3480
TTAGAAAATA TGT	3493

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1299 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

CTTCTTTGGA GATTAGGACT TTCTCAATAT CTTTTTCTAT CATAGTGCGC TCCTTTCAGC	60
ATTACTCCTT TGTCTTAAAA AGGAGTCTGT AGTGTATTTT ATCAGTTTCT AAAGGAATAC	120
TCAAATAGGA ATTTGCAAAT TTTGGTACCC AAATGATTTC CTCTTGCTCA TCCGTAATGA	180
CCCAAGCACT CTCTCTCGCT TCATTTGGCA CTTTTTGATC AATAAAAACT CTGCTTAATT	240
TTTTCGTAAA CCCTGGTTTT AATGTAATCC GATCACCACCT TTGGCGATGA CGAACGGTCA	300
AAGGTGTGGC TGTGGTCAAC GGGATCGCTA ACAACTGACC CGTCCACTGA TTAATCTCTT	360
CAGGAAGAGG AAACGGTTTG TCTGTCGCA TTAATCCTAG CCACTCGGTT TCCGATAGAT	420
AAATCCCCGC ACCAGGTAAT AATAAATGTG TAACTTCTTC TCGCAGTGCA GCCTGTTTTT	480
GTTCTAAGCA TAGCCAATCA TAACTTTTAG CTAGTTGCCA TTGCTGTTCT AACATAACCG	540
AGCCTTGCGG TTGCCGTTCA TCATTCAAAA CTTTTAAAAT TTGTGTCATC TGGCGTTGGT	600
TTAACGAAAC CCCTAATGGC ACTAATGTTT TTTGGaAAAA CGCGGTAAAA AAGAAGTGCT	660
GGTACGCAGG CGTTTCTTTT AATAATCTCT TCAAAGGAAT GCGCCAGCCT TGTTTGTTG	720
GTTCTACGGC TTCAGATAAT TGAGGCTCAA TTTGTTCTTG AATAAATTCT TGGGCAAAAC	780
GTTGTTCAAT ACTAAAAGAA GCAATCTGAT CTAAGAATTG CGGATTTTCC TGCTTTAATA	840
GCGGCAAAAC CTGATTACGG AGCCGATTTC TCAATATTC ATTGGTTTGG TTCGTTTCGT	900
CCTCAAAATA AACGAATGCT TGGCGCTGTG CAAATTGATA CAATCTTCT TTTGGATAAA	960
TGAGTAACGG ACGAATCAAT TCCCCTGTG CAAAAGGACG TCTCTCTTA ATGCCTGCGG	1020
AATGGGAAAA ATTAGTTCCT CGAATCAGCT TCATTAAAAT CGTTTCCGCT TGATCGTCTA	1080
GATGATGTGC CGTCATTAAG ACCGCCGCAC CTTCAATTTC CATGACTTCT TtAAAAAAt	1140
CATAACGAAA TGTGCGCGCT TTGCTTCTAA ATTtCGTTGC TTTcAgGGTC TcCCAACGTG	1200
TTcATAATAC GTaGcgCTCG CTGCTGACAA TAGGTAGCGA GATATTGGCG CTTCTGTGC	1260
CGACGCTTCT CGCAATTGAA GGATTGACAT GTGGCAAnA	1299

## (2) INFORMATION FOR SEQ ID NO: 149:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4503 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

TTACTAGTCT GTTGAATAC ATTATGATAG TGTAAGATAA ACGGACTATT GGAAAGTCAC	60
AGCAGAGGAG TTTGAACATG TACACATTTA AAATCGGTAT TCCGGCAGAA GTACACGATA	120

CATTTGTTAA	AAATCACCCA	TTATGTAATT	TACTACAGTC	ATCTTCATGG	GCAAAAGTCA	180
AGGATAACTG	GGGTTCTGAA	ATTGTTGGGG	TTTATGAAAA	GGATACATTA	GTTGCATCAA	240
GTTTAGTTTT	AATCAAACCA	TTACCAGCCG	GTTTTACGAT	GTTATATACA	CCAAGAGGTC	300
CAGTAATGGA	CTATACAAAT	GAACGTTTGG	TCAGCTACTT	TATGGCTGAA	TTGAAAAAAT	360
TTGGCAAGAA	AAAACGAGCT	TTATTTATAA	AAATGGATCC	AGCGGTTTCAT	TATCAAGATT	420
TTCATTTGGG	AGAAGAACAC	CAACCGCATG	CTGAAGCGAC	ATCTATTATT	GAGACCTTGA	480
AAGAAGCAGG	TGCCAAATAT	CAAGGACTGA	CAATGGATAT	GGGGGCCACC	ATCCAGCCTC	540
GCTTCCAAGC	GAATATTTAT	CGTGAAGATT	TTAGTGAGGA	ACAATTGTCA	AAAAGTACCA	600
AGAAAATGAT	TAAACAGCT	GAAAAAAAAG	GGGTCGTTGT	GCAGCAAGGG	CATGTTGACT	660
TTGTTGATGA	ATTTGAAAAA	GTCATTCAAT	CAACCATGGA	ACGCCAACAT	ATTTCTTTGC	720
GGAATAGTGA	TTACTTTAAG	AAGCTGTTGA	ACATTTACCC	AGAAGACTCC	TTTATTATGC	780
TGGCGCAAGT	CAATTTAAAG	GAACGCTTAG	ATAGCACGCG	ACAACGCTAT	GACAAAAATC	840
AAAAAGACTT	GAGTAATTTG	AAAGAAAATC	AAGTGAAAAA	GCGCCATAAT	TTAGAAGAAT	900
TAGATGCTTC	CTTGACTCGG	GAACATAAAG	AATTAGAAGA	AAATATCGCA	GAAGCTGGTG	960
AAATCGTCAC	GGTTGCTGGC	GCTTTGGCTG	TGACATTTGG	TCCAACAAGT	GAAATCCTTT	1020
ATGCTGGCTT	AGATGATCGT	TATAAACGCT	ACATGCCAAC	ATATGTGACT	TGGCGAGATG	1080
CCATTCAAGA	ATGTTTTAAC	CGTGGCTGCG	AAACTGTAAT	ATGGGCGGCT	TAGAAGGTAG	1140
TTTAAATGAT	GGCyTAATTA	AGTTTAAAGC	GAACCTCAAC	CCmACGATTA	ATGAATTTAT	1200
TGGGGAATTC	GATTTACCGG	TTAATAAATT	ATTATTTTAA	GCCAGTGAAT	ATGCCTATAA	1260
ATTACGAAAA	CAAAAAAAT	AAATAAAAAA	CACAGGGACT	AAAAATCAAA	CTTGATTTTT	1320
AGTCCCTGTG	TTTTTTTCTA	TAGTTTTTGC	GTTGCATTCC	GTTTCATTAA	GCGAACAAAA	1380
CGCTAGCAAT	TGCTGAAGCA	GCTACACCAG	CAACAAGAGC	TTCTGGCACA	CCAGAACTGA	1440
AAATCGTCGC	AATAATAGCT	AGGTAAACAG	CAGTTCACAG	TTACCAATA	ATTTCAGCAT	1500
AAGCTGAATT	AAAAAGAAAA	TAAATCATGT	TCATGACTAA	AATGGTATTC	GTTGCTGAGC	1560
CTAAAACACC	AGCCACAAAT	AAGGAGACAG	GTTGCGCTTT	TTCCTTCGTT	AATTTATTTA	1620
GCCATTTGTA	AGCGAAATAA	GGAACGACCC	CAATTAAAT	CCGCGGAATG	AGCACAAATTA	1680
AAAGTGCCTT	CCAACTACCA	TGATCGGTAC	CGATCACAGG	AATAAATGGA	GAAAAAACAA	1740
ACGACAAAGG	TGTTGGTGCT	GTCGTATTTT	TAATTAAGCT	AATGATCCCA	AACGTGCCGC	1800
CCAACATACT	ACCGATTTTT	GGCCCTAAAA	CAATGGAACC	AATAATTACA	GGAATATGCA	1860
TCGTGGTCGA	ATTTAAAGGA	CCTAATGTGA	TAAATCCTAG	TGGGGTTAAA	GCCAGTAAAA	1920
GTAAGATAGC	TAAGAACATT	GCTGTTAACG	TAAATTTTTT	CGTGTTATTC	ATAAATTACT	1980
CCTTGTTATC	AGATGCTAAA	ATTTTCATCA	CTGTTTGAAG	AATATCTTCA	ACAGTGGCAA	2040
GTGCGCCACG	ACCATAATCG	CCACAAGCTA	ATAGAGCTTC	ACGTGGGATT	ATTTCTTGAT	2100

AGCCGACTTC	TTTTAAAATG	CCAATATTC	GTTGGACAAT	GGGGTTTTGG	TACATATACG	2160
TATTCATCGC	GGGCGCAATA	ATTTTTGGTG	TTTCAGGATA	AAGTGCCAAA	GCAACTGTTG	2220
ATAATAAGTC	GTCGGCAATC	CCATGGGCTA	ATTTCCCAAT	TGTATTGGCC	GAAGCAGGTG	2280
CTACTAAAAA	TAAATCCGCT	TTTTTAGCTA	ACTCAATGTG	ATTAATTTTA	CTGGGATCAA	2340
TTTCCATCAT	GACATCCGTA	TGCACTGGAT	TCTTAGAAAG	TGATTGTAAG	GTCAGCGGAG	2400
TGATGaACGC	TGTGCTACTT	TTAGTCATTA	AAATTTCCAC	ATTGTAGCCA	AGCTTAGCTA	2460
GTTGACTTGT	AATATCTGCT	GCTTTATATG	CAGAAATGCT	GCCAGAGACG	CCTAATAAAA	2520
TAGTTTTTCAT	TGTTGTTCTC	TCCATTTCTT	TTCAACATTT	TTAACGATCA	TTGAACCAAT	2580
CCCTTGTTTT	GAATAAGCTT	CTTGGACAAT	GCCCGTTTCG	TTAATTAATA	ATCCTTTGTG	2640
TTCTGTTTCA	TGAACGTTCA	TCAAGTCATT	GGCTAAGACA	AAGTCACACC	GATTTTTAAC	2700
CAAAGCGGCT	TGTGCAACTT	GGACAAGTTC	TTCTTGTAAG	ACGTTGACAA	GTAATTTAAA	2760
GCCCACTAAA	ACAGTTTGGG	GTTGCTTGTC	ACGAATCATT	TGGATGACTT	TAGGATTTTT	2820
CTTTAGAAAA	ATTAATAATC	GGTCCGTATC	AGATGAAATT	TTGTTGCTG	TTTGAGGAAT	2880
ATCAGCAATC	TGATTCAATG	TGTTTTGAAC	GATAGTTACC	AAGGCTTCTT	TGGTCTTTGG	2940
CAATGTTTGT	TCGCTTAATT	GTTGAGCAAA	aGAATCAATA	AATTGTTCTT	CCGTTTGCGC	3000
GGTTTCCGTC	GTAATAATCAC	TGACTGCCAT	TGCGTGAACA	ATCGCATCAA	ATTGCTTTTC	3060
AGCAAAGAGT	TGTTCTAACG	TAGTGGCTAA	TTCAACCGTT	GTTTCTATTT	CTTTAATTGA	3120
TAAGTCCAAT	TGTTGGGTTG	GACGTAATGC	ATGTTTGGTT	GTTACGTAGG	TTACGGTGTG	3180
TCCAGCAGCT	AAAAaACTTT	CAGCAATTGC	TTTTCTTAGT	CCGCCAGAAG	AATGATTGGT	3240
AATCGAACGA	ACATTATCAA	TCGGTTCTGA	CGTTCCGCCA	GCAGTAACTA	AAACATCCAT	3300
AATCAAAACG	CCTCCTTAAG	AAATTCGGAC	AATGGTCCGT	CCAATATGTT	GGCCTGCTTG	3360
TAATGCTTCA	AAAACAGGGG	GAAGTTCTTC	TAAAGTAATC	TCTTGATACA	CAGCTTTTTC	3420
AGTGATATCT	AAATCAGTGG	CAAGACGTTG	CCAGACTATT	TTTCGTTTGT	CCATACCAAC	3480
ATTaACCGAA	TCAATGCCCA	ATwAATGAAC	GCCCCGTAAA	ATAAAGGGTA	AAACGGTTGT	3540
CTCTAATGTA	ATTCCAGCAG	CGCGGCCACA	AATGGCTGAA	CGGCCATCAT	AACGAAGTTG	3600
TGGTAAAACA	GCAGAAGTAA	TTTCCCCACC	AGTGGTGTCA	ATTGCAAAAG	CAAAACGCTG	3660
TTTCATCAAT	GCTTTAACAG	GAGTTGCTAA	GAAATCTTCT	ACTAATAAGC	ATTCATAAC	3720
ACCTAACTCT	TTTAAAGCT	CAAAAGAATC	TGGCTTTCTA	GTTAATGCAG	TGATTGAAGT	3780
GTAGCCTAAT	TGATGcAACA	TCGCAwTAGA	CaATGTCGCA	ACaCCGCCTG	TCGCACCACT	3840
TACAATAATG	GGGGCTTCTT	TGTCGTCGCG	TAAGCCaTCG	TCTTCTAAAG	CTGTGACAGA	3900
TAATCCGGCC	GTAAGCCAG	CAGTTCCGAA	GATCATTGCT	TCTTTGAGAC	TCATTTTTTC	3960
AGGAAGTGGA	ACTACCCAAT	CGCCAGGAAC	ACGCACATAT	TCAGAATAAC	CACCAGTATG	4020
ACTAACCCCT	AATTGATAGC	CTGTTACGAT	CACTTTATCG	CCTTCTTTGA	AaCGAGGATC	4080

862

TTCCGATGTG ACAACAGTGC CACTGACATC AATTCCTGGA ATCATTGGAT AATTGCGAAT	4140
CACACCGCCG TCTTTTTTTG AAGCTAAAGA GTCTTTAAAG TTAAGTGGG AGTAGGCTGC	4200
TTTGATTACT ACAGTGCCTT CTGATAAATC GTCTAAAGTA ATGGTTTCTA GTTGACCAAT	4260
AACAGATGTA TCTTGTTCTC TTAACGTAAA TGCTTGAAAT GAATTCATAT AAAAAGTCTC	4320
CTTCCAATTA TGTAAGAAAG CAACTCGTCT TATTTTATCA TATTCTGATT TTTTCGTATC	4380
ATTTTCTTTC TTATAAACAG TAAGTTATAC AGAAAAGAAA GATTGTGTGA CAATTAGACT	4440
AGAAAAATAA AGAAAAATGG TTGAAATTTT TCAAAAAGTT CGTATACTAA AAAAGGTAAA	4500
ACT	4503

## (2) INFORMATION FOR SEQ ID NO: 150:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11739 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

GCGTTAGGTC AAGTAAGTCT TGGmTCGTTT TTGATGAcGT CTTTTTAtGG TGTTTTTmaG	60
TACTmGTTTA TCGTATTTTC TATTACGCAA AAATCAATCC AATTATTTT TATTATTGAT	120
GATTGGCATG ATTTTAGGTA CTTTTTTCAG TAGTATTAGT ACGTTTTTAC AAGTGATTAT	180
GGATCCTAAC GAATACGACT TACTCCAAGG AAAGTTGTTT GCCAGTTTTC GTAACGTCAA	240
TAGTCAGTAT CTCATTTTCG TTGGTGTGGC GGTACTCATA ATTTGTGGGA TACTTTGGCG	300
GGAAAGTTAC CGTTTAGACG TTTTACATTT AGGGAACCAT CaAGCACAGA ATTTGGGGAT	360
TGCTGTGAGT CAGTTTCAAT TAGTGTGTG GTGTTTGATC AGTAGTCTGG TCGGCTTATC	420
AACAGCATTG GTGGGACCGA TTACATTTT AGGTTTTATC GTGGCCAATA TGAGTTATCA	480
ATTCATGAAA ACCTATCGCC ACCGAGAACT GTTTGTGGCA GGGAGTTTAA TAGCAGTCTT	540
TTTACTGGTG TTTGGTCAAT TGATGGTTGA ACAGGTCTTT CATTTGAATA CCCCCTCAG	600
TGTGGTGATT CAATTTGTTG GGGGCGTTTA TTTTATATGG AAAATTATTG CAGAAAGGAA	660
GCAACGCACA TGATCCAGAT GAGTGATGTA TCAAAAAAAT ATGGCGATAA ATTTGTCGTT	720
TCAGACATTG ATTTACCTAT TTCTGAAGGC AAATTAACGG CTTTATTGG ACCAAATGGC	780
GCTGGGAAAA GCACGCTGCT AGCCATGATG AGTCGCTTAA TTCCCAAAGA TACCGGGGAA	840
ATTTATCTGG ACAAGCAAGA AGTAAAAACC TGGAAGCAAT CGGCTTTTTC ACAAAAAATT	900
GCCCTTTTAA AACAAGCCAA CGGTGTACAA CTAAATTAA CTGTGCGAGA ACTGGTCAAT	960
TTTGGTCGTT TTCCTTATTC CAAAGGTCGC TTAAGGAGCG CTGACCACGA AAAAGTAACA	1020
GAAGCACTGG AACAAATGGG CTTAACTGAA ATGGCGGAAG AATACATTGA TACCTTTTCT	1080
GGGGGCCAAT TACAACGGGC GTATATTGCC ATGATTTTAG CCCAAGATAC CGATTATATT	1140
TTATTGGATG AGCCCTTAAA TAACCTCGAC ATGAATTATG CGGTTCAAAT GATGCAGATT	1200

CTTAAACGCT TAGTGGATGA ACTAGGAAAA ACGATTCTGA TTGTTCTACA TGATATTAAT	1260
TTTGCCGCAA GCTATGCAGA TGAAATTGTC GCCATGAAAG GCGGAAAGCT CTATGCGCAT	1320
GGCGCAACAG AAGAAGTTAT TCAAACAAGT ATTTTAAATG ATCTTTATGA AATGAATATC	1380
CGGATTTGTG AAATTGAAGG AAAACGCTTT TGTCTTTATC ACTAAATCGG GCAATTTAAA	1440
AATAACGTTT TTGGAGGAAA AGAATGAAAA AGAAATTTT AGCAATGATG GCAGTTTCAA	1500
TGATGGGACT GTTAATGTTA AGTGCTTGTC AAACAAATAA AAAACAGCA GATTCTGCAA	1560
CAACAGAAAC AACAGCTAAA ACGGAAGTCA CAGTCAAAGA CACCAATGGT CAATTAACCG	1620
TTCCCAAAAA TCCTAAGAAA GTCGTTGTTT TTGATAATGG TTCCTTGGAT ACAATGGATG	1680
CACTAGGTGT CGGTGACCGC GTGGTAGGTG CGCCAATAA AAATATCCCT GCGTATTTGA	1740
AAAAATACCA AAAAGTTGAA TCAGCAGGCG GCATTAAAGA ACCAGATTTA GAAAAATCA	1800
ATCAACTAAA ACCAGACTTA ATTATTATTT CTGGTCGTCA ACAAGATTAT CAAGAACAAT	1860
TAAAAGCCAT TGCGCCAACC ATTTACTTAG CTGTAGATGC CAAAAATCCT TGGGCATCAA	1920
CGAAACAAAA TATCGAAACG TTAGGCACTA TTTTGTATAA AGAAGAGGTA GCTAAAGAAA	1980
AAATAACTGG CTTAGAAAAA GAAATTGCTG ACGTGAAAAA ACAAGCAGAA GCTAGCGCGA	2040
ATAATGCGCT TGTTGTGTTA GTTAACGAAG GACAACTTTC CGCTTACGGA AAAGGCTCTC	2100
GTTCGGTTT AATTCATGAT ACATTTGGCT TCAAAGCAGC AGACGATAAG ATTGAAGCTT	2160
CCACTCATGG GCAAAGTGT TCTTACGAAT ATGTTTTAGA AAAAAATCCT GGGATTCTCT	2220
TTGTGGTAGA TCGCACCAAA GCAATTGGTG GCGACGATTC AAAAGATAAC GTCGCTGCAA	2280
ACGAATTGAT TCAAAAAACC GATGCTGGTA AAAATGATAA AGTCATTATG CTTCAACCAG	2340
ATGTTTGGTA TCTAAGCGGT GGTGGATTAG AATCAATGCA TTTGATGATA GAAGATGTTA	2400
AAAAAGGATT AGAGTAAAAA AAAGCCTAGG AAAAAATCA CTTTGGATT TGTCTCTAGG	2460
CTTAAAAGCT GATAACGGC GGGAACAAAA GCAACTCCTT CGGAAATAAG CTGAAATTCT	2520
CCAAAAATTA AAGAACAATT TTCGGAAAT TCTTCTTAT TCTCGGAGCT AAACACTTCT	2580
GTCCCGACCT TCTTGATGAA AAACAAAAA GTTTTAAATT AAAAATGTGG TTGACGGCCT	2640
CGAAAATGTT TGCTAAACTA TCCGTAAATC AGTATGATTG ATGAAACAAA TAAAAGAACA	2700
AAATGTGATG AGAAGGAAAG TACTTTTTAA AATAGCTTGC AGAGAGCCTC GTTTGGTGAG	2760
AAGAGGTAGT GACGTTAAAG GGGAAATGG CCTTTGAACA GGTGGTTGA AGCCGTTTGC	2820
TTAAACCAAT TCAGGTGCGC CTGTTATAGC GCTACAGTAT GATTGTACTG GAAACGAGCG	2880
GATCTTTTTT AGCGAAGAGA TCAACAAAGG TGGTACCGCG ATTGCATAAG CATTCTCCT	2940
TTTTAGGACG AATGCTTTTT TTATTTGTTT AAAAAATATG AGGGGGCAAC AACATGAAAT	3000
TGGGAAAAA AGTAGTAGGT TTGATTGCAA CAGGGTTTCT TTTAGCCGCA TGTGGCGGAA	3060
CCAAAGAAGC GGCAGAGAAA GTAGATTCGG GAAATTTAGC AGCTGAACAA AAAATCAGTA	3120
TTAGTTCACC TGCACCAATC TCAACATTGG ATACAACACA AACAACAGAT AAAAATACCT	3180

TTACAATGGC	ACAACATTTA	TTTGAAGGCC	TTTATCGGTT	TGATGATGAT	AGTGCCACGG	3240
TGCCAGCTCT	AGCTAAAGAT	GTCAAGATTA	GTGACGATGG	GCGCAAGTAC	CACTTTACCT	3300
TGCGGGAGGG	GATTAAGTGG	AGCAACGGCG	AGCCAATCAC	GGCCCAAGAT	TTTGTATTATT	3360
CTTGGAAGAAA	ACTGGTGACA	CCAGCGACGA	TTGGACCGAA	TGCCTATTTA	CTAGACAGTG	3420
TTAAAAATAG	TTTTGAAATA	CGCAACGGTG	AAAAGTCAGT	CGATGAATTA	GGGATTTTCAG	3480
CCCCGAATGA	CAAAGAATTC	ATTGTTGAAT	TAAAACAGGC	CCAACCTTCC	TTCTTAGCAG	3540
TCGTTTCGAT	TGCTTGGTTA	GCGCCACAAA	ATCAAAAATT	TGTCGAAGCG	CAAGGCCAAAG	3600
ATTACGCCTT	GGATAGTGAA	CATTTACTTT	ATAGCGGGCC	ATTTACGCTA	GCCAATTGGG	3660
ATGCGACTTC	AGATACTTGG	ACATTGAAAA	AAAATCCAGA	ATACTATGAT	GCGGATCAAG	3720
TGAAACTGGA	AGAAGTTGCG	GTTAGCACAA	TCAAAGAAGA	TAATACTGGG	ATTAACCTTAT	3780
ATCAAGTGAA	TGAAGTAGAC	TTAGTTCGCA	TTAACGGACA	ATATGTTCAA	CAATATCAAG	3840
ATGATCCAGG	CTATGTCAGT	CATCCAGATG	TGGCCAACTA	CTTCTTAGAT	TTCAACAAAA	3900
AAGAAGGAAC	GCCATTAGCG	AATGTTTATT	TACGAAAAGC	GATTGGCCAA	GCAATTGATA	3960
AAGAAGCCTT	AACACAAAGT	GTCTTAAACG	ATGGGTCAAA	ACCCCTTAAC	GGATTGATTC	4020
CAAGTAAACT	TTATGCGAAT	CCAGAAACGG	ATGAAGATTT	CCGAGCTTAC	AGTGGCGAAT	4080
ATTTGAAAAA	TGACGTCAAA	AAAGCTCAAG	CTGAATGGAC	GAAAGCCCAA	GCGGATGTCTG	4140
GTAAGAAAGT	GAAACTTTCA	TTGCTGCGG	CAGACACAGA	TCAAGGAAAA	CGAATTGCTG	4200
AATATGTTCA	AAGTCAGTTG	CAAGAAAATC	TGCCAGGTTT	AGAAATTACC	ATTCATCGC	4260
AACCAAGTAA	TAATGTGAAC	CAATCGCGAC	GTGAAAAAAA	TTATGAGTTG	TCTCTTTCAG	4320
GATGGATTGC	CGGCAGTAGT	GAATTAGACT	CTTACTTTAA	CTTATATGCA	GGAGAATCAA	4380
GTTACAATTA	CGGCAATTAT	CATAATGCCA	AATACGACCA	ATTGGTAGAA	GAGGCACGAA	4440
CGATTAATGC	CAATAATCCA	GAGAAACAGT	TTGCAGAATA	CAAAGAAGCG	GAAGACATCT	4500
TGTTGAACCA	AGATGCTGCC	CAAGTACCGC	TGTATCAAAG	TGCCTCAAAT	TATCTAATCA	4560
ATCCTAAATT	GAAAGGCATT	AGTTATCACT	TGTATGGGGA	TTATTTCCAC	TTGCGCAATG	4620
CCTATTTAAC	AGAATGATTC	AGGAAGAAGG	AGTAAAAAAT	GGAAAAATTA	GTCGaACGAT	4680
TTTTAACTTA	CGTGAAAGTT	AACACCCGTT	CAGATGCCAA	CAGTCAAACC	ACACCGACGA	4740
CGGTCGGTCA	AGTGGTTTTA	GCCAAAATGA	TTGAAACAGA	ACTTCATGAA	TTAGGCTTGG	4800
CCGATGTGCA	TTATAACGAA	CAAAACGGCT	TTTTAACGGC	ACGGTTGCCT	GGTAACCAAC	4860
CTGCAGCGAA	GAGTATCGGC	TTGATTGCAC	ATCTGGATAC	GCGGATTTT	TCAGCAGAAA	4920
ACATCCGTCC	CCAAGTCATC	ACTAATTATG	ATGGTCAAGA	TGTCCTCTTG	AATCAAGAAC	4980
AAGGCATCGT	TTTATCAGTT	GCTGAATTTT	CCAATTTTAA	AGAGTATCAA	GGGGAAACGT	5040
TGATTACTAC	AGACGGAACC	ACACTTTTAG	GAGCAGATGA	CAAAGCGGGA	ATTGTCGAGA	5100
TATTAGCTGC	TGTTGAGTAT	TTTTTGCCAC	ATCCAGAAGT	AGCGCGGGGG	GATGTTTGGC	5160



TCGCTTCGGT CCAGATGAAG AGATTGGACG CGGGGCCGAT CAATTTGATG CGCCGAACTT	5220
CCCAGTCGCT TTTGCGTATA CGATTGACAG TGGCCGCGTG GGGCATTTCG AGTATGAAAC	5280
GTTTAACGCT GCCCAAGCTG TGATAACAAT TGAGGGAACC AGTGTCCATC CGGGAACGGC	5340
CTATGGTTCG TTAGTAAATG CCATTAAATT AGGCGAACAA CTTGATCAAT CACTACCGCA	5400
AAAGGAAGTG CCAGAACAAA CACGTGGAAA TGAAGGATTT TATTTACTCA ATAAATTTAC	5460
AGGAAGTATT GAAAAAGCAG AACTTGTCTA TATCATTCGG GATCATGATC AAGAAAACTT	5520
TCAAGCGAGA AAGCAATTTT TGGAAAAACA AGTCCAACGT TTAAATGCGC TGGCGGATAA	5580
ACCCAGACTA ACAATTACTT TCCAAGACCA ATACTATAAT ATGAAGGAAA TCATCGAAAA	5640
AGATTGGACT CCTGTAGAAT TGGCCGTCCA AGCAATGGAA AGTTGTGACA TTGAACCAAT	5700
CATCACACCA TTTCGTGGCG GGACAGATGG TTCTAAGATT TCTTTCATGG GCATTCCCAC	5760
GCCTAATCTA TTTACAGGTG GCGAGAATTT TCATGGGCAA TATGAATTCA TTACCGTTGA	5820
ATCAATGGCC AAAGCTGTTT AGACGATCAT TGCCATCATT CGCTTAAACG CTAATTAAAG	5880
AAACACGCAA GGGAAATGCC GATTTTCCTT GCGTGTTTCG CTGTTTTTCG GTAAAGTATG	5940
TAGAAAAGTA GGGAGTTTTC CTCATGAATT ATCCaTTGAT TTTTGAGAA AACCAACGAA	6000
TCGAAACCGA ACGCTTGATT TTAAGGCCCC TCACATTAGC CGATGCAGAG GATATGTACG	6060
AATATGCTTC AGATGAGGAA ACGGTCCGCT ATGTGTTTTT GAAAAATCAA ACAATCGCTG	6120
AAACAAGACA AAACATTGCT AAATTTTTTA TGGGGGAACC TTTAGGAAAA TATGGGATTG	6180
AAAGTAAAGA AACGGGCAAA ATGATTGGCA CGATTGATTT GCGGGTCAAT GAAACCAATA	6240
ATATTGGAGA ATTAGGCTAT GTCTTGAACC GTACCTTTTG GGGCAATGGC TACATGCCTG	6300
AAGCCGCGAC GGCCTTAGTT GAACTGGGT TTGCAAAAAT GAAACTCATG AGAATCTTTG	6360
CTTTACATGA TCAAGACAAT CCAGCTTCTG GTCGTGTCAT GGAGAAAATT GGTTTTACCT	6420
ACGAAGGAAC ACTACCCAAT GCTCGAATTA GTAAAGGGAA AATTGTCACA GATGTTTATC	6480
GTGGGATGAC ACTGGAAACA TGGCAAAATC GACAAAAATA GAAGCGGAAG ACAAGTTAAG	6540
AGAGGAATTA GAGAAATGTT GATGTTTACA GAAAAAGAAT TCGCAGCCTT TGAAGTTGCT	6600
GGATTAGATG AACGAATGGC GGTATCCGT GCGCAAATTC AACCTATTTT TCAAGAATTA	6660
GACACGTATT TTGCCGAACA ACTAGCGCCT GAACTAGGAA CGGAGTTGTT TGTCCATATT	6720
GCGCAACATC GCCGTGCGAC CGTGTATCCG CCAGAAAATA CTTGGTCAGC ACTTAGTCCG	6780
AATAAGCGCG GGTATAAAAT GCAGCCGCAT TTTCAATTAG GCATTTGGGG CGACTATGTG	6840
TTTATGTGGT TGTCATTTCAT TGATAATCCT AAAAATGAAA AGCAGATTGC GCAACCTTTT	6900
TAGAGAATCA GCAATTATTC CAAGCATTAC CAGAAGATAC GTATGTyTCC TTGGATCATA	6960
CGGTTCCTCA AATTACTCCG CTGCCAGAAA CGGATTTAGA GAAAGCCTTA ACTCGTTTTA	7020
GAGATGTAAA AAAAGGTGAG TTTGAAATTG GGCGCATCAT TCCCAAAGAT AGTGACTTGT	7080
GGCAAAACCC TGAAAAAGCA CGAGCTTATA TGTTGGCAAC CTATCAACAA CTCCTGCCAT	7140

TGTATCAATT	AGCTGTTGCC	CAATAAGAGA	AACAGACAGT	TATTTTGTGT	TTTAGTGTTA	7200
TTTTATCGAA	TTATCCAATC	GTCAAAAATG	AGAAaAGCTC	ATAATGTTAG	TTTAAATTAC	7260
TAACATTGTG	AGCTTTTCTG	TAGTTTTTTT	GAAAAAAATT	AAGGGATGAC	TGGCTGATTT	7320
CTATGGAAAA	GTACGTTTTG	ATAGCCTTTT	TTGCTTAAAT	AACAAGAGTA	GCAACCCTTG	7380
TTTAATAAGG	GATTTTCGCG	AATAGATGTT	GTCAAATGAA	AAAAAATGAT	GTTTTGAACA	7440
AAGAGTTCTT	TGCTATTATA	AAGGTATGAA	AAAGGTTGTT	TACGCATTCT	TTGTATAGAA	7500
GAATGTGAAT	GAGTTAAAAG	GAGGTTTTTA	AATGAAAAAA	AAGCGTTATT	TAATAATTGC	7560
GTGTTTACTA	TTTTCCCCTA	GTTTTTTTAT	AAATGTTGAA	GCATCTGAGG	GTGGTTCTAG	7620
TTCGGTGGA	ATTGAATTTT	ACCAAAATCC	GGCAACACCC	GCTCCTAAAG	ATGCCCCACC	7680
GAAACAGAT	GAGCCAGCTG	CGGATCCCAA	GGAACCAGCT	GGTCTCTGC	AAGGAGATCA	7740
ACGAaGTGGT	GGTTCGACAC	AGACCACCAC	AGCTGGCTCG	CAGCTCCCTC	GTACAGGAAG	7800
CAAGAGTCAG	GCAAACCTGA	GCATTCTTGG	TCTTGTCTTG	ATTGGTCTTG	TCGGAATGGT	7860
CCAGAGAAAAG	AAGGGACGAC	ATGAAGCAAA	CTAAGTGGA	ACGATTAGCA	ACCATTGGCT	7920
TGTGTAGTTC	TTTAGTGATT	AACGCCTTTT	CTGGCGCGAC	CGCAGTTGCG	GAAACAATCA	7980
CGATTGAAAG	TAGTCCGACA	GTAGCAAGTA	GCGCTAAGGA	AGCAACATCA	GCAAGTAGCG	8040
CAACGCCAGA	AAGTACGGAA	AGCAGTCAAG	AAACGACTGA	AACGAGTCGG	GAAGAAGTGA	8100
CACAGGAAAC	AGTGGAACAA	GAAGATACAA	CGGAAATAGC	TCAAGAAGAA	AATCTGCTGG	8160
AACCATCTGG	TATTCAAGCG	AAAATTATGC	CTAGAGCATT	TGTCTGGGCA	GCTACAGGAA	8220
CCACGAATAG	AACCACGGTT	CGTCTTAACG	ATGTAGTTGA	TTATCGAATG	AAACTGGAAA	8280
ATATTACAAC	AGATGATACA	AATTCAGATA	TTGTTGGTCC	GATTCTGAAT	GAATCAAAAT	8340
TGGCAGCTGG	AATGACTAGA	CCAACGTCAG	TTATTATCAA	AGTTGGCCCC	GAGGAACAAA	8400
ATGTAGTTAT	AGGAGAAGGC	TCTCATAATG	CCAATAGTGG	TGGCGAATAT	TTTGTTTGGA	8460
CAGAGTCGAC	TAGGACGGTC	ACTGCATTTA	TCAATCGTCT	ACACGGGCGA	ACTTCTGGAA	8520
ATACTGGGTA	TATCAAAACA	TTGTATTTTC	AGACAAGAAT	TACTTCGGGT	ACACACAACG	8580
AAAAAAAATA	TATCGATTCC	ACGATTTCGAT	TTAATGGATT	TGGCGCAGTG	AATAGACGAA	8640
ATGAAATGAC	GTATGTGGAT	AGATTTTCTG	GTAACCTAGG	TTTGAAAGGG	AATTTAAAT	8700
TTTACGATGA	GGATGGAGTG	GCTAGTAAGG	AACAAATGAG	TCTTTCCATA	AACCCAGTAT	8760
TATTTAAATC	AACCAATTCG	AGAGTAGAAG	GGACGCTGTT	AAGATATCCT	TTAACTATGA	8820
CAAATGCTGG	TAGTGGGCTT	TATAGGTTTG	ATACAGGACT	AAAAAATATA	GGTACTAGTA	8880
ACGGGGGTTC	GTGGTTTTTC	TTTAGTAACG	ATATTGCGAT	TAAwTTTGCT	ACAGCTACCA	8940
CCAATAGAAT	TGATAAAATT	GTGTTAGCAC	CACCTAAAGC	AACACTTCCA	GACAGCCGAA	9000
CAGGTGTGAT	GACTTCCAGA	GCGTACACAT	CAACAACCTA	TACGAGAAGT	AACTATGATA	9060
TTTCaCaTAT	AaCTTTACTA	AAGCGACTGG	TAAGACATGG	CGTATAACGA	ATCGTAGTTC	9120

TAATAGCGGA GTAAGCACCA ATATTATGTT TAGTAACAAT CTTGTAGGAA CAACAGGACA	9180
GTTCAATAAA GGGAAATTCTG GTGCTACTAT TAATTATTAT ATTTATTACA AAAAGCTTTA	9240
TGAGAATTTT GTTAATAGTA ACGGACAAAA AATTACACCA CCATCAGGAT TCACACAAGG	9300
AAAGCGAACG GTTATTAATA GTGAAGCGTA CACTTTTAAA CAAAGTGGCA CCTTGCCAGA	9360
TACCTATCAA GCAGACGGCA AAACCTATAA GTTCAAAGGT TGGTACAAAG GCAAAACCAA	9420
GCCAAACACC TTGACCACCA CCAAAGCACC AAGCTATGCG GTGACTTATA ACAATGATGA	9480
TGATTTGACG GTGGTGTATG AGGAGATTGA GTTTTTTGAC TTCCCAGCGC TGACCTATCA	9540
ATTGCGGTTT GTGGACGAGT CTGGCAAACG AGTAGATGCT TCAACAATTG GGTGACTTA	9600
TGACAAATGG CGTGGGGAGT TGTTAACAAA TAATGTCAAT GACTGGAAAA CAGTCAGTCT	9660
TGAAAAAGGT CAAGTTGCAC CAACCAAAAA CAATTTGAAG GAAATTGCTT ATCCTGCGCA	9720
ATCTCTTGAA ATACTGAGTG ATCGAAGTAC ACAATATAGT GCGGCTAATT TAACCTTTAC	9780
ACTGCCCAAA TACTATGGGA AAATCAGTGT TTATAACAAG AGTGGTACTT TTGATGCAGG	9840
CTACCTGTTT CCAAATATCC AAAATAATAA TAGTTCAGAA CCATCTGTTG AAGGCTTACA	9900
CAAAAATCTT TCTAGCTGGT TTGAACTAAA GAAAAGTGGG AACCATCGT TTGTTTTTAA	9960
CAAACCGATT CCAGCAGCAC CAATCGATGT TCAGGTGCCT GCTTACTTGC GAGAGATTGT	10020
CTATAATCCC GATCTGATAT CAAATACAGC CGTCTATTAT GTAATAGATA AACCTGTCTA	10080
CTACTATTTA ACCAACCGCA AAGTCACCGA AAACCTCGTA GACACAAGCG GTGCCAAAAT	10140
TACACCACCA ACAGGTTTCA CCCAAGGAAA ACAACAGCG ATTACCAGTG ACCCATACAC	10200
CTTCAAACAA GCAGGTACCT TACCTGATAC CTACACAACA GGCGGTAAGA CCTACAAGTT	10260
CAAAGGCTGG TACAAAGGGA AAACCAAAAC CGAGCCTTTA ACCACGACGA AAGCACCAAG	10320
TTATGGGGTA ACCTATAATG ATAATGATGA TTTGAATGTG GTGTATGAAG AAGTAACAGC	10380
TATTTATCCA GGGTCTATG CGCAATTTGT TGATGAACAA GGCAAAGCTT TTAATAATTC	10440
GCTAACTCTT TCTGAAAATT ATACTGAGTT TCTCAGAAAG ACAGGCACTA CTGTTTTTGA	10500
AACGAAGGGC GCCTTTTACC CAATGGTAGG TAAGCGCGAT GCAACAGTGA CAAATAAGTA	10560
TAAGATTGAA ACAACGAGTA GTGTTGCAGT ACCAAATAAT TATTGGACAG CCATGCCTGC	10620
TGACATTGCT AATAGAGGGT ACGCATTAAT TAACCTGGAT ATTACCAATC AAATGAAATA	10680
TGTTGACAAG ATAGAACTA CTGACACGCA CACTACAATG TACGATTATA CTGGAGTTTC	10740
GGACATTGGG GGCAAACCAA ACAATGCTTC TCAGACAGAA TTAAATCCTA TTGATAAAAC	10800
ATTTGTTAAG CAAACAGATA ACACTTTTGC GATAAAATCA ATTTGGGGAA ATGGaACGTT	10860
TTCGCATATA GTTCCCAAAA TGTTATTGTT AGCAGAAAGT GGCAGTTTTT CAAGATTATT	10920
CGGGTTTGAT GGTACTCCAG ACTACAAGCA AACCCTAAAC TACAAAGTCA CCCGTAAACA	10980
AGTCACCGAA AACTTTGTTG ATGCCAACGG TGCCAAAATC ACCGCACCAA CAGGCTTCAC	11040
CCAAGGCAAC AAAATACCCA TGAAGTACAA CACCTTCAAG TACACAGCCG CAAAAGCCTT	11100

GCCAGCAACT TATACCGCAG GTGGCAAGAC CTATACTTTC CAAGGCTGGT ACAAAGGCAA 11160  
AACGAAGCCT AATACTTTAA CAACCAGCAC TACGCCAACA TATAACACAA CTTTTGATGG 11220  
CAATGACGAC ATGACGGCGA TTTATAAGGA AGAAGTGCCC AAAGCAAGTG TCGCTTTAAC 11280  
GCGCACTACC GCTGAAACCG TGA CTAGCGG CGGCAATGTC ACCTGGCGCG CCACAATTAC 11340  
CAATACAAGC CAAGCACCAC TAACCACGGC GACCATCAAG AAATCAACCG CTTGGACAAC 11400  
TGGCTTAGCA GCACCAACCG CCATGATTGT CACGCCAGCT GCGGGAACCG CAAAAACAGT 11460  
TCCCGTCACA GCGACTACGT GGACCAACGG TGTTAGCTTA GGCCTGACA TTCCCGTAGG 11520  
AAAATCTGCC ACGGTTCACT TCACTACGAA AGCGACGGGG ACAGCGGGGC AAGTCTTGCG 11580  
CGCTGGGATT ACCACCAAGT GGAATTACAG TGGTGTCTT ACTTCGGCCA CCGTACGTGT 11640  
GAAAGACAAT GATCAAGCAA TTGTAACACC AACGGCTGAA GGCTTCATCA GTGTTCCAAC 11700  
CTTCAATTTT GGTCAAGTTG GGGTAGCGGG AAGCACGCA 11739

(2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4045 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

GAATTTCTTC AGTATCCTAA AACAAGAAGT GTATTATGGA AAAATCTATC AAaGTCAAaT 60  
GAATTGrTaG AAGCAaTTGA GAACTACATT TATTATTACA ATCATCACCG AATCAAGGAA 120  
AAACTTAACT GGAaAGCCC AGTAGAATTT CGACAATTCA ATCAaAAaAC TGCATAaAAA 180  
TAGAGTGGAa AAATCCACTC TATAaAAAGT CTAaCTTTTT GGGGTCGCTA CAACAATACG 240  
TTTGTTTTTT AATCCTTCAG TTTATAACAT TATAAGAGAA AAGCGAGAAG AATCATCCTC 300  
CTTCTCGCTT TCTTGTTACC CTAAACCAAC ACGTTCAaAA ATCGTGTCaA CATTTTTTAA 360  
ATGATAATGG TAATCAaAGG CATCATCTAG CTCTTCTTTT GATAAGACAG ATGTAATTTT 420  
TTCATCCGCA TCTAGCAACG GACGGAACGC CGTTTGATGA TCCCATGCGT AAGCTGTTTT 480  
TGGtTGGACC AGATCATACG CTTCTTCACG TGTCATGCCA TGATCAATCA ATTTTAACAA 540  
GACACGTTGG CTGTAAATTA GACCATAGGT CGCATCCATG TTACGTTTCA TATTTTCTGG 600  
GAAGACCGTT AAGTTTTTGA CAATATTACC AAAACGATTC AACATGTAAT TTAGTAAaAT 660  
CGTCGTATCT GGAATAATAA TCCGTTcAGC AGACGAGTGA GAAATATCTC GTTCATGCCA 720  
TAAACTCACA TTTTCAaAGG CAGTAATGAC ATGACCACGA ATCACACGTG CCAaACCAAGT 780  
CATATTTTCA GAACCAATTG GATTCCGTTT ATGCGGCATT GCCGAAGATC CTTTTTGTC 840  
TTTTGCAaAA AATTCTTCTA CTTACGTGT TTCTGATTTT TGTAAGACCAC GAATTTCTGT 900  
TGCAaATTTT TCAaTACTTG TCGCAATTAA TGCCATGGcA GAAACATATT CTGCATGTAA 960

ATCACGTGGT	AAAAC TTGTG	TCGAAATCTC	TTGAGCACGG	ATCCCCAAAT	GTTTCGCAAAC	1020
ATATTETTGA	AEAAATGGTG	rAATATTGGc	AAAGGTACCT	ACGGCTCCAC	TGATTTTACC	1080
CGCTTCGACA	CCTTTGGCTG	CATGTTCAAA	ACGTTCAATG	TTGCGTTTCA	TTTCAGAATA	1140
CCACAACGCT	AATTTCAAAC	CAAATGTAGT	CGGTTCTGCA	TGAACGCCAT	GCGTTCTTCC	1200
CATCATGAmC	GTATGTTTAT	ACTCTTTGGC	TTTGGCTCCG	ACAATCTCCG	TAAATGTTTG	1260
TAAATCTTTC	CGTAAATAT	CATTGCGCTG	TTTCAATAGA	TAACCATAGG	CAGTATCCAC	1320
AACATCTGTG	CTGGTTAACC	CATAGTGTAC	CCATTTTCGC	TCTTCACCAA	GCGTTTCTGA	1380
AACCGCACGA	GTAAGCGGA	CAACATCATG	ACGTGTTTCT	GCTTCAATTT	CTAAATACG	1440
TGCAACATCA	AATGATGCAT	TTTCTCGGAT	TTTTTGTACA	TCTTCTTTAG	GAATTCGCC	1500
AAGTTCTGCC	CAAGCTTCAT	CCGCCAAAT	TTCAACTTCT	AGCCATGCCT	TGTAGCGGTT	1560
CTTATCTGTC	CAAATTGCGC	CCATTTCGGG	TCTTGTATAA	CGATCAATCA	TTGTGCCTTA	1620
TCCCTCTTT	ACTCTCAAT	TTTGAAAAA	TAATTACGAC	CAAATCTTTG	TTTCTTTAAT	1680
TTCTTCCAGT	GTTTCATCGA	TTTCTTTGGt	TAATATAGTA	ATATGcCCCA	TCGCTCGTTG	1740
CGCTTTTGCT	TCwCTGTGGC	CATAATAGTG	GAAATGCCAA	TCTGGATGTT	CATTAATAAT	1800
GTCCATCGTT	TCATAGGCTT	CATTTCCAAG	CACATTGACC	GTGACTGCTT	CAGACAACAG	1860
TTGAATCGTT	GGCATTGGCC	AATTACATAA	ACCACGAATA	TGCGTATCAT	ATTGGCTAAA	1920
AGAACAAGCA	TCTATTGAAA	AATTCCCAGC	ATCTTCTGGA	CGAGGTAAAA	TTTTATTTAC	1980
ATAAATCCCG	CCTGTTTCTG	TATAAAACAT	TTCAATTGAA	AGGGTGCCGA	CTAAACCAAC	2040
TGCTTTAGCA	ATCACCTTAG	CAATTCGTTG	CAACTCACCA	CTGAcTtCGT	CATCAATTCTG	2100
GGCTGGTGCA	ATCGTTTCAT	GTAGcTTACG	CTCTTTTTTA	ATTGTTTCCA	CAACTGGAAA	2160
TGTCGCAAAT	TCGCCACGAC	CATTACCAGA	AATCAGAACA	GATAATTCTT	TTTCGTAAGG	2220
AATCCAAGCT	TCCAAGACAC	AGGTCCTTC	TCTCAGAAGA	TTCATGGATG	GCGCTAAGTC	2280
TGAAGTACTA	TGTAGTACAT	AACGACCTTC	CCCTCGCGTA	TTCTTCAAGA	TACACGGATA	2340
ACCAATGCCA	TCAATTGCTT	CTTGAATATC	AGTCGGACTA	ACAATCGTTG	CATAAGGTGC	2400
AATCACAATA	TTATTATCTT	CCAAGAAACT	TTTTTCCATT	AAGCGATCTT	GCGTAATCGC	2460
TAAAATGTCT	GTTCTTGCG	GTATATTCGC	CAAATGTTGA	ATCGCTGTTA	ACGCATCAAC	2520
ATCTGCATGT	TCAAATTCAT	ACGTAATCAC	TTGGGCACGT	TTTGCCAATT	TCTCTAACGC	2580
TTCAGGATCA	TTATATGCTC	CTAATAATTG	CCAATCGGCA	ACTTGTGCAG	CAGGACAATC	2640
AATTTGTGGA	TCTAGCACAC	CCACTCGAAA	CCCCATTCT	CTTGCACTAA	GGGCCATCAT	2700
TTTTGCTAAA	TGTCCGCCAC	CAATGATACC	AATCACTTGA	CCAGGTAATA	GTGGTTTATC	2760
CAAGTTCATC	ACAACTTTCC	ATCACAGTTT	TATCCTTAAA	ATCCTTCTGC	GGTTTCTCGT	2820
CGCAGTCTAG	CCTTTAAGAG	TGGCCCTnTT	TACATGaTAC	mAGGrAAAAG	GcTCAACTTC	2880
AAGTGCCAGC	AGTGTTCCTA	TTAGGTTTTT	TCATCAGAAT	TTATTAACAT	ACTTTAGGaT	2940

AAAACAGCGG TAACTTACTA CTCGTCTCTT T6TAACCAAG CTTTTTCTC TTTTCTCTT	3000
CAAATTGATG AAAACATTAA ATTTTCTAA ACGATGATGA GGTCAAAAAA GTATGAAAGA	3060
AAAGCCACCT GACCTTATCT ATCATACTTT TTTGAAACTA TTCGTTACTT TCTGAGGTTT	3120
CTTTTAAACC ATTGAATAAA AGATTTAAAA TAATTGCGGT AATGCTGCTC ATGACAATTC	3180
CATTACCAGT AAACATGCTA ACAGTTTCTG GTACTTTATT AAAGAGCGTC GGCATAATAT	3240
TAAAGCCTAA ACCAAAGCCA ATCGAGATGG CAATAATTAA TAAATTTTGT TCATTTGAGT	3300
AGTCAACTTT TGACAACATG CGCATTCTT GGACAGCCAC CATTCCAAAC ATAACCAGCA	3360
TTCCGCCGCC TAAAACCGGC TCTGGGATAA TTTGGGCCAT CGCACCAATT TTTGGTAATA	3420
AGCCTAAAAT AATTAAGAAA AAAGCAGAAA AATAAATTGG TCGACGTGTT TTGATCCCAG	3480
ATAATTGAAC TAACCCGACA TTTTGAGAGA AACCTGTATA AGGAAACGTA TAAAAATCC	3540
CCCCTAAAAT GACGGCTAAA CCTTCTGCAT GATACCCACG ACGTAAATCA TCTTCGCCAA	3600
TTTTCTTCCC TGTAATATCT CCTAAGGCAA AATAAACACC TGTTGATTCA ACCaTACTAA	3660
CAATtGaAAT AATAATCATC AACACAATAG AAGAAAGATC AAACGTTGGT TTGCCAAAAT	3720
AAAATGGCTG TgGAAAtGa aACCACGTAG CTTGACCAAC AGGAGACAAa TCCACTAATC	3780
CCAGAAACGC AGCTAAAATT GTGCCACCGA CTAGGCCGAT TAACACGGCA ATTGACTTGA	3840
TAAAGCCACG TCCCCAACT TGCACAACAA TGATTAAGGC AATTGTTACA AATGCGAGGA	3900
GCAAGTTTTT AGGATCACCA AAACTTTTGT CCGTTGCTAA GCCGCCACCC ATTTTTTCAA	3960
CAGCAACAGG AATTAAGGTC AAACCAATAA CTGTAATGAC TGTGCCGGTG ACTAAAGGCG	4020
GAAATAGACG TTTAATTTTT GAGAA	4045

## (2) INFORMATION FOR SEQ ID NO: 152:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7574 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

nAnnAATATA CTATAAGTC CAAAGAAAGC GCTTTAATTC GATGGAGAAG GTGGTTTAGT	60
GAAAAGATTT TCATTTTTTT TACTAATTTT ACTTGCTTTA ACAGGTTGTA AATCCGGTGA	120
AAAAGAATTT GATGAAGAAT CTCTTCAAAA TCTAAAGGAA ACGkCACAGT CTTAyTCAGA	180
AACAGAATTA CAAAATGGTG ACGTTCGTTT AAATGAATAT ATTTCTTTGA AAGGGGAGAT	240
TGTTGAGAGT GACAGTCGTT CCAGTTTAAT AAAAAAGGT GATCGTTTTA TTTTGAAAAG	300
TGGTTCTAGT AAATATCAAG TTTyTAATGA GCAAAAGAAA AAATTGAAGA TTGGTGACGA	360
AGTGACAGTT TACGGAGAAT ATTACGGCTT TTTGAAAGGG ACATTAATTG AAAGTGAGGA	420
GAATCATGAT TCAGCCACGA ATTAGACAGA ATCCTTTTGA AACATACGAA GAAGTTGCAG	480
AAGGATTTGA AGATTTATTA GCGCCACTGG AATTGTTTTT TGATCGAGAG TATCAAGGTC	540

ATCTGGATTT AGGGACACAC GGAACGGTTT ATTCTAAAGG GACACGTGAT GCGGAGGCAT	600
TTTTACGACC ATTATGGGGA CTAGGTCCTT ATGTAACGCA AAATGAAAGT GAGTATTTGA	660
ATGACTTTCT AACGGGAATT ATCGAAGGAA CCGATCCTGA GAGTTCATCT TATTGGGGaA	720
AAACAAAAGA TTATGATCAA TTGATTGTTG AGATGGcAGC ATTAAGTACA TTTCTTTTgt	780
TAAACAAAGA AAAAACGTGG GACCAATTaA CAAAAGAACA GCAAAACAAT TTACACAGTT	840
GGCTGATACA AGCCAATGAA AACATTATTC CGCCAAATAA TTGGCATTTC TTTAGAGTGT	900
TGCTGAtTac GtGAACAtgT GAAgCCGTAT cnAAAGaCa GAtCGCTGTT GATTTGGCTG	960
TAATnGATCG TTTTtAtGtC GGTaATGGTt GGTATTATGA CGGAGTCGAG ACACAAGTTG	1020
ATTATTATGT TTCTTtCGCT ATTCATTATT ATAGTTTATT GTATTGTCGT TTTGCCCCGG	1080
AAGATACCGC AAGAGTGGCT ATTATGAAGG AACGAGCAAC TTTGTTTCGCG CAAGAGTTTA	1140
AATATTGGTT TACTCAGCCA GGAGAAGCGA TTCCCTTTGG ACGTAGCTTA ACGTATCGTT	1200
TTGCTCAAGT GTCATTTTTT AGTGCTTTAG TTTTTCAGA TGTTGAAGCA TTGCCGTGGG	1260
GTGAGATTAA AGGTCTAATC AGCCGACATC TTCATCAATG GATGAACAAA GATATTTTTA	1320
CAACAGATGG ATTATTGTCG GTAGGTTATG ATTATCAAAA TATGGTCTTT GCCGAAGnTA	1380
TaATGGTCCT GGTTACCTT ATTGGGCCTT tAAAACATTC ATTTtGTTAG CGGTTCCAAA	1440
AGATCATCCT TATTGGCAgC AGAAaCACaG CcATTAGCT TtCCaGAAAA ACACCTACCT	1500
cACCgAAAGT CGGAACtACT ATCAaGTAAA TGATGCAGGG ACACATGGTT TAATGTTTCC	1560
AGCGGGACAA TTTATTAATT ATCAAGCTCa TGCGCATGAT AAATATAGTA AATTGTTTA	1620
TTCTTCCCAT TTTGGTTTCA GTACTATAAA GAGTGATTAT TGGTATTACG AAGGTGCTTA	1680
TGATAATTGC TTGGCTTTAG CTGAAGACGA CCACTATTTC CGCACGAAAG GATTAGATGA	1740
TCAGTACGAA ATTCTTGATG ATCGTATCAT TCATCAATGG CATCCTTGGT CCGATGTAGC	1800
AATCAAAACG ACAATTGTAC CATTAGAAGG CCAACATTTA CGGATCCACG AAATTGAGAC	1860
ACAACGTGCA TTAGTGGCTT ATGAAGGTGG CTTCTCAATC CCGCTCTTCG ATGAAAAGGT	1920
CACTTGTGTT TCTGATCAAA TGGCTGAGGT AAAAAATGCA AAAGGTGTGT CAAAAGTAGA	1980
AAATATCAaC GGCTTTTCTG AAGCyGCGAT TATTGAACA GAACCGAATA CAAATTTACT	2040
TTATCCGTTA ACGGAGTTGC CATATTTGAA AGCAAACCTG TCGAAAGGGA AACATCTTTT	2100
AATTTcATTG GTGACAGGCG TACTACCGAA TGAACAAATT GAGCCAGTTA AAGTTAGGTT	2160
AAAAGAAAAT CAGCTTTTAG TTGAAGAAAA AGTTGTaATT TTGGGGAATT AGGGGGAGAA	2220
ATACACAATG ACAGAAAATG GACAGATTGA GTGGTTGCCG AAGCAAATCA GTTTTGCTTT	2280
AAACAAAGTT AGAGGAAATT TAGAGACATT TAAAGAATTA GTTCCACCCG CGGCAAGTTT	2340
AAATCAAATT TATTATCCTG AAGAAAaTAT TGACTGGACG GCCAGCTTTT GGCCAGGAAT	2400
GTTATTTTTA GCAAAGGAAT TAACAAATTC AACAGAGTTC GATGAGGTGA TTGCCACACA	2460
GATGGCTTCT TTTCAACATC GCTTGGATGA ACAAATTGAG CTGGAAACAC ATGATATAGG	2520

CTTTTGTAT	ATTTTAACAG	CCATTGCTGA	TTATCAAGTC	AATGGTCATG	AGGCCTCAAA	2580
AGAAATGGCC	TTGCAAGCTG	CGGACTTATT	AATGAAACGT	TATTCACCAA	AAGCCAAGAT	2640
TATTCAAGCT	TGGGGGAATT	TAGACGATCC	AGAGCAACGA	GGTAGAATGA	TTATTGATTG	2700
TTTAATGAAC	TTACCATTAT	TATACTTTGC	AGCAAAAATG	ACTGGCAAGC	AAGAGTATTA	2760
TGAAGCTGCC	TACAATCATG	CGAAACAAAC	ACAAAAATAT	ATTGTGCGAG	AAAATCACAC	2820
TACGTTCCAT	ACGTATTATT	TTGATACGGA	AACAGGCGAA	CTTGTACGGA	AAGACACAGC	2880
AAGGATACTC	AGATGATTCT	TGTTGGGCTC	GCGGCCAAGC	CTGGGGAATT	TACGGCTTTA	2940
CGTTAAGTTA	TTTATATACA	GGAGATAGTA	GTTTTCTGGA	AACAGCAAAA	AATGTTGCTG	3000
ATTACTTTAT	TCmAGAATTA	CCAGAGGrCa	AAATCTGTTA	CTGGGATTTA	ATTTTCAATG	3060
AAGGGAGTGA	AGAAGAACGT	GATAGCTCAT	CTGCTGCAAT	TGCCGCATGT	GGCTTACTAG	3120
AATTATCTCG	TCAGCTTCCA	TTAAATGATG	AAAAACATGG	CTATTATGAA	AAGGTTGCGC	3180
TAGAACTATT	ACAAGCATT	GCGGAAAAGT	ATACAACGGT	TTTACAACCA	GAGTCGAACG	3240
GCCTATTGCT	TCATGGTGTC	TATGATAAAA	AAACAAATAC	AGGAGTCGAT	GAATGCATGA	3300
TTTGGGGCGA	CTACTTTTAT	TTGGAAGCGT	TAACTCGTTT	AGCGAAAAGC	TGGTATTCTT	3360
TTTGGTAATT	CATGGAGGAA	AACTCCTAAA	AAATAATAAA	TAGCGGAGGT	AGTACAAATG	3420
AAATTTTGA	AAAAAGGCTT	AACAGCGGCA	GCGCTGTTAG	CAGTGGCGGC	AGTAACTTTA	3480
ACAGCATGTG	GTGGTTCAAG	TGAAAAGAAA	GCAACTGAAA	AGAGTGAAGA	TGGCAAAACA	3540
AAATTAACAG	TAATACTT	GAATTATGAC	ACGACCCAG	AATTTGAGAA	ATTATTGAGA	3600
GCTTTTGAA	CGGAAAATCC	TGATATCACT	ATTGAACCGG	TGGACATTGC	TTCAGATGAT	3660
TATGACACAA	AAGTAACAAC	GATGCTTTCA	TCAGGAGATA	CGACGGATAT	TTTAACCATG	3720
AAAACTTAC	TTTCATATTC	TAATTAcgCG	CTACGCAATC	AATTGGTGGA	TTTAACCGAT	3780
CACGTTAAAG	ATTTAGATAT	CGAACCTGCC	AAAGCAAGTT	ACGAGATGTA	TGAAATCGAT	3840
GGTAAACCT	ATGCTCAGCC	TTACCGTACA	GATTTCTGGG	TATTGTATTA	CAATAAAAAA	3900
ATGTTTGATG	AAGCCGGAAT	TGCCTATCCC	GATAACTTAA	CTTGGGATGA	ATATGAAGCG	3960
TTAgCGAAAA	AATTATCTAA	ACCAGAAGAA	CAAGTATATG	GTGCCTATCA	ACATACTTGG	4020
CGCTCAACCG	TTCAAGCGAT	TGCTGCTGCT	CAAAACAATG	CCAATTTGAT	TGAACCAAAA	4080
TACAATTATA	TGGAACTTA	TTATGATCGC	GCATTGAGAA	TGCAAAAAGA	TCAATCACAA	4140
ATGGATTTTG	GAACAGCAAA	ATCAACAAAA	GTAACGTATC	AATCACAATT	TGAAAATTCA	4200
AAAGCGGCGA	TGATGTACAT	GGGTAGCTGG	TACATGGGGA	CTTTATTAAC	AAACATTGAT	4260
GATGGCAAAA	CAAATGTCGA	ATGGGGGATT	GCCGAAATAC	CACAACAAGA	AAAAGGCAAA	4320
GCAACTACCT	TTGGCTCACC	GACAAGTTTT	GCAATTAATA	AAAACAGTAA	AAAACAAAAA	4380
GCTGCTCAAA	AATTCTTAGA	CTTTGCTTCA	GGTAAAGAAG	GTGCAAAACT	TTTAGCAGAA	4440
GTAGGGGTGG	TTCTTCTTA	TAAACAGAT	GAAATTGATA	AAATCTACTT	TGCAAGAAAA	4500



GGAAaTGCCTT CAGACGAGTC TCACAAAAAG CCTTTAACCC AGATACAATT AATTTAGAAT	4560
TCCCATCTGA TAAAAACGGT GCCGCAATTG ATAAAGTATT ACAAGAmGmA CATGATTTAA	4620
TTATGGTCGG CGmCGAAAAA CCAAAAGATG GTATTGCTAA CATGGAAAAA CGTGTCAAAG	4680
CAGAGATAGA CTAATTATTA TTTTCTGCAA AAACACAAAA GATTGATGAA AAAAATAAGC	4740
TGAAATGAGT AACCTACAAT GAATTTTCAGC TTATTTTTTT AAATAAAGAA TGTAGCTTGA	4800
GTTTGGAAGG AGGGCCTGTT AAGTTCACTT AACAGAATAC AATTATGGAA ATTACTTCAA	4860
AAACAAAAAC AGCAAATAAA TTGAGACGAA AAAATACGTG GACTGCGTTA TCTTTTATCG	4920
CACCAAACCTT TATTGGCTTT TTCCTCTTCA CTTTAATCCC GGTTATCTGT TCACTGATTT	4980
TAGCTTTTAT GAGTTGGGAT TCTTTTCAA CCCCTGAATT TGTcGGAATG AAAAATTCA	5040
CCAAAATGAT TCATGATGAT ACGTTTGGGA TTTCGTAAAA ACAAACGTTT ATCTATACGA	5100
TTGGTGTTGT TCCGTAAACA TTGATTGCT CATTAGGACT AGCGATTTTG TTGAATCGTA	5160
AAATTAGAGG GATGAAATTT TTTAGAACGG CTTTTTCTT TCCTTATGTA ACGTCATTAG	5220
TTGCGATTGC GGTCGTTTGG AATATGCTAT TTCATCCAAC GATGGGACCA ATTAATCAAT	5280
TTTTGAAATT ATTTATTGAA AATCCTCCTG GCTGGACATC TAGTTCTACT TGGGCTTTAC	5340
CAGCAATTAT CATCGTTAGT GTGTGGCGCT TTATGGGGTA TTATATGATT TTGTATTTAG	5400
CTGGTTTACA AAGTGTCCTA AGAGAAATAT ATGAAGCAGC GTCAATGGAT GGCGCGGGTA	5460
AATGGAATCA ATTTTAAAT GTTACCTTAC CTTCACTACG ACCAACGACC TTTTTCGTGA	5520
CTATCATGCT TGTAATTAAC TGTTTCAAAG TGTTTGACTT AGTTCAAGT ATGACTGGTG	5580
GTGGACCAGG ACGTGCAACG AATGTGTTAG TTTACGAAAT TTACAATGAA GCCTTTGTCA	5640
AATTTAATTT TGGTTATGCA TCAGCCATTG CGATGGTCTT ATTCATCATC GTTCTAGCGA	5700
TTACTGTGGC CCAATTTAAA TGGAATCAAC ATCAAGAGAA AGTTTAGGAG GAGAAAAGAT	5760
GGAAGCAACT TTTTCAAAAA AAGAAAATAC TCGAAAAGCA AAAACGaAAA AAGCAGAAAA	5820
AGTCACACCT GGTGGGATCC TTGTAATCAT TCTCTTAACT TTACTTGCGT TAATTACTTt	5880
GGTGCCTTTC ATTTGGATGG TTTCAGCATC ATTTAAAnCh AACAAATGAAG TTTTCACCAT	5940
TCCAATTCAA TGGATTCCTA AAAGTTGGCA TCCAGAAAAC TATTCCGTTA TTTGGGAACG	6000
CATTCCGTG TTAACCTTCT TTAAAAATAC ATTATTTTTA AGTATTGTGA TTACGATTAT	6060
TCAACTATTT ACTTCAAGT TTGCAGCATA TGGCTTTTCA AAAATGAATT TTAGAGGAAG	6120
AGATACACTT TTTATTACGT ACATCGCAAC AATTGCTATT CCTTGGCAAT CGTATATGAT	6180
TCCACAATTT ATTATGATGC GTCAATTAGG TTTGACTGaC ACGTTATGGT CACTTGTTTT	6240
aTTACAAGCA TTTAATGCAT TTGGGGTTTT TTTACTGAAA CAATATTACA GCAGTATTCC	6300
AGATTCACTA TGTGAAAGCG CGCGGATTGA TGGGTAAAGT GAGTGGGGAA TTTATTGGAA	6360
AATTATTTTA CCATTAACCTA AACCAGCTTT AGCTaGTCTC ACGATTATTA CTTTkGTGAA	6420
TACATGGAAT GACtATATGG GaCCTTTCAT TTATTTATCT TCTACTGAGA ATAAAACGAT	6480

TCAGTTAGGT TTGAAAATGT TCGTCGGTTT ATTTGATGCT GAATATGCCT TAATCATGGC	6540
CGCATCAGTG GTTCTATTT TACCAGTAGG GATTGTCTTT TTAGCAATGC AAAAATACTT	6600
TGTTGAAGGT ATTGCTACAT CTGGAATGAA GAATTAGGAG GAAGTTTGAA TGTTTACTAA	6660
GCAAACCTTT GATAATAATA TCTATATGAA AATTTTTTCGC TGGCTTTACA TTTTATTTAT	6720
CGGTAATCTA GGGCTTCTCT TAGTGAACAT TCCTTTTTTC ATAGCGGTGA TTAGCTTAGA	6780
CATTGATCCA CGAAACCTCC CACTTTTTGT AGTAACGTTG TTGCCGATGG GGACAGGGAT	6840
GATTGCACTT TTAGGTTTAA TCGATACTTT TAAAGAAGAA AAGGAGCTGG ATCCATTTAA	6900
AACTTTTTTC CAAAAGTTTC GTCAGTTTGG GTTAAGAGGC TTTCTTATCA GCTTGCTTGG	6960
ATTAGGAAGT AGCATCATCA GTGTGACAGA TATATTCTTT TTTGCCAAAA CAACGATTGG	7020
TAAATGGTTT ATTCCATTGA TGGTCCTTCT GCTAATTTTT GGGTTAGCTA TTATGCTGAA	7080
TGCTTGATAT TTTCAAGTAC GAAACCCGCA AGCTTCTTTT AAAGATGTGT TTCGGATTTT	7140
TGTTTACTAT GGACTGCGGA AATGGTATGT CAGCTGTCTA AATGTTTTTT TACTATTCTC	7200
AATGTTTGCA ATGATGTTTT TGAAGCCACA GTTTGGGTTT GTTGTAACCC CTGTCCTGTT	7260
CCTAGGGATT ATTTATTTAA ATACTGGGAA ATTGCATGAA AAGCAAAAAA AGAATAAATA	7320
ATTTTTACAT AATAAGTAAT CTCGATTTTC AGGAAAATGA ATGAAGGTGG GGTGCTTAT	7380
TTTGTTCTTG CCTTGTTTAT TAAAGTGGAG GGAGCGAACG AATGAAGAAA AAAATAATTA	7440
TTTTAAGTAG CTTTTTGATA GGTGCTATTT TTGTGCTCGT TATTTTTACT ACCAGCCAGC	7500
AAAGTATGAT TAATAACgAa GCmATCCAAG TCgArGAATT AATTGTTwAT GGCGATAAAA	7560
ATATGAATCG TATA	7574

## (2) INFORMATION FOR SEQ ID NO: 153:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1229 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

TGTnTAATGT TCGTGTTTTA TnTTTTAATA ACCTTTTTTCG AAAAAATATA CGAACGTTTCG	60
GGTTTAAAAA CGAACAAAAA GATGAAAATG TGTTGAAATA TACGTTTTCC TTGCGTATAC	120
TGATGAAGTG AAAAAACGAA CGAGGAGTGT ACGAATGAAC AAGAAAGTTC CTTATTTAAT	180
TGTGGCACCG GGGCTAGTAC TATTATTATT TTTCTTAATG ATTCCATTAA TTACTAGCAT	240
TTTGCCAACC ATTTTTACTG ATCATGGCTT GACACTAAAT CAGTATGTTA CTTTTTTTAA	300
AGATGACTAT AATGTTTCAA TTTTCTGGCG AACGATAAGA GTTCGTTAA TTGTGACGGG	360
TATTTCAATC GTGTTGGGCA TTCCAACGGC ATACTTTATT GCTGGAGTTT CCAAAAAATG	420
GCGTGGCTTT TTAATGGCTA TGACCTTGTT TCCATTATTA ACCAATTCAG TTATTCGAAG	480

TTTTGCTTGG ATTAACATTT TAGGAAAAAA CGGTGTAGTG AATACGTTGT TATTAAAAAC	540
AGGCTTAATT GAGCAACCGC TGAATTTATT GTATACAGAA TTTGCAATTA TTATCGGTTC	600
AGTGTATTTA TTTTACCAA CAATGATTAT GACCTTAGTT GGGGTATGG AAAACATTGA	660
AGGAGAAATG CTAGAAGCGG CTGAGACGCT AGGCGCCAAT CCAATGACGG CTTTTCGGAA	720
AATTGTTTTA CCTTTATCAA TTCCAGGCAC CATTGTAGGA AGTATATTAG TGTTTACAGG	780
AACTTTAACG GCTTATACGA CGCCACAGCT GTTAGGTGGT AACCAAAAAA TGATGATGTC	840
GACCTTCTTA TATCAAAAGG CTAATACTTT AGGTGATTGG CAATCGGCGA GTGTTCTTGC	900
ATTTATTATG ATTTTAACGA CATTAAATTGT GATGAAAGGC CTAGACATGG TCGCTAAGAA	960
AGTAGATCGG AGAGAAGCAA ATsATGCGTA AsAAAAAGGA CTATCTGTTA TTGCAATTCT	1020
AGTTTTTsTG TTTTtATTTT TACCGTTATG TTTGATTGTG GTGcnTCTTT tGGAaCAGCC	1080
GCAGCGATTG AATTTCGAT TAAAGGATTA CTTTAGATTG GTACGCCAAA GCATTACAAT	1140
CGGAAACGTT TATGGACAGT TTAAGTTAAG CTTGTTGAAT GGGGTATAGC GACTGTTTTA	1200
GCCATTAATA GTAGGGATCC GGCCTCCnA	1229

## (2) INFORMATION FOR SEQ ID NO: 154:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9749 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

CTTGGTTATC TGGTATGGCC ATCCTCTGTG TGAGCAAAAA TTATGGAGAT CTATTnGTAA	60
TATTATTTGT GTGACTCCGT ACCCTCTCCT TGCTGTGGCT GATAAGGGCA TCTATGTGGT	120
TGAAGCACGA AAGrAGGgtT mCGTGTATTg rTaATCCTGT AGcgTATTaA CaATwACCCa	180
cATTTGaTTa CCaATTATTT AATTTAAACA ATTATCGTGT ACTTTCAACT AGAACTCCaA	240
AAAATAATTT TTCAGATCAA aTAGAGTTAG ATATTTATAG TAGAGGAATG GGTGGTATTG	300
GGTTGCCAGG AGATTTATCA TCAGTATCTA GATTTGTAAA AGCAACTTTT ACTAAGTTAA	360
ATTCTGTATC AAGAAGTTCA GAATATGAAA GTATTAGCCA ATTTTTCAT ATTTTAAGTT	420
CTGTGCAACA ACAAAAAGGA TTGTGTGATG TTGGTGATGA AAAATATGAG TATACGATTT	480
ATTCTTCATG TTGTAACCTG GAAAAGGGAA TTTATTACTA TCGTACGTAT GACAATAGTC	540
AAATTACTGC TGTGGATATG AATAAGGAAA ATTTAGAGAA GGATAGCTTA ATTGTTTATC	600
CAATGGTGGa AACACAACAA ATTAACATG CTAATTAATT TGTGTGAATT CTTGTTTTTA	660
TAAGTATTTT CGAAATGAGT TTAAATAAAC GGGaATAATA GAaGAGTAGA TTTaAGTTAT	720
TGATAACTAC TAGGTAATTG ATGATAAATT AAmCAAATTA GATGCTTTAA GGAAAGATGA	780
CAAAAAAGTA CAGTATTACA TGAAAAAGTG GATTTTGTAG AAAAAAAGGG TTTACAATT	840

AATATCAAAA ACAACTATTC TTATAAACG AATAGTTGTT TTTTGTATT CATATTTTCT	900
ATAAAAAAAT GAATAGTAAA CACTTGTAAT TTTTAAATAA GAAgAATAGC TTGTTATAAG.	960
AGTAACTACG ATTGTCCGCA TAATTTTTTt GAAGgTGgTA AAAAAAAtAT AGrGkTATAA	1020
ATGtGaTAGG kGtATTAGtA tTTATTTAaa AGrCtAAAAA AtGtGaAGGA GkTTGaATCg	1080
AAaTGaAAAA AGTATGtGCT ATTTTAATGA TTTCTGCGTC AGTTTTAGGA ATGGTGCCCA	1140
GTAATGTACC AATTTTTTACA GCAGAAATAA ATTATGCGCA TGAAATGAA TTACAAATGA	1200
ATCCGTTTGT TATAGCGTTT GTCGATAAAG TGAAGGGTTC TGTTGCAAAG TTTGAAAATC	1260
AAACGCGCCC TCTCTGAAct CCAAATATCT TAGGCTGGTA TTCCCATTA CTTGATTTTC	1320
AGTAGACACC GAAATCCGA AGAGCGTTCC ATTTCTTCGG TTCTTTTTAT ATATTCTCTG	1380
AATGGTCTCC ATGCCCTTAA TCGTGGAAGA GGCTGTACGG AGACTTTGAT AAAATTTATT	1440
CCGTCGTTTA ATAGGTCGAT GGTCTTGTTT TATTAAATTG TTAAGATACT TCACAGTTCTG	1500
GTGCTCTGTC TTAGTATATA AACCCACACT CTGTAACTTT CTAAAGGCGG AGCAAGaGAA	1560
GGTGCTTTAT CGGTCACAAT TGCTTGCGTA TCCCGTTTCT TTCGTAACCA GATATCTAAG	1620
GTTAAGCCGT CCGCATCAAT TGCACGATAA AGATAATGCC AACGTCCCTT AATTTTGATA	1680
TAGGTTTCGT CCATTTTCCA TGAATAGAAG GATTGTCTAT TTTTCTTCTT CCAAAGATCA	1740
TAGAGGACTT TGCTGTACTC TTGCACCCAA CGATAAATCG TAGTATGACA AACATTTATT	1800
CCACGATCAT ATAACAATTC CTGaACTTCA CGATAGCTTA GATTGTAACG CaGGTAGTAA	1860
CCAACAGCGA CAATAATGAC GTCTTTTTTG AATTGTTTGC CTTTAAAATG ATTCATTACT	1920
CTGTCCTCTC TGTCTTTTTT CTCAATTTTA CACTAAAATA GATTTTTTGG AAAACTTTGC	1980
AACAGAACCT TTTTTTGCAA TAGAATTACT AATGGGAAAA CAAATTAATG CGATAACCAG	2040
ACCGATAGCA ATTTTCGTGAA AAATAATTGT TATTAAGAAT AAACTATAC TAGATATGAA	2100
TAACTTAA ACATGTATTT TATCTTCTTT CATAGGATCC TCCCGTTATT GTCATACAAT	2160
AGTTGTTTGA TTTATATTTT AAATTATCAA TAAAAGAAAA GCTAGGAAAA AAGCATTCAA	2220
ACGAAAACCC AGTTAATACA GATACAGGCG TTGTGATTGT CGCAGTTGAA GATAGTAAAC	2280
CGATTATTCA ATTAGCAGAC GGAACAACAA AGAAAGTTGA AGCTAAAGAA ATTGGAGCAG	2340
ACGTTCAAAA AGACGGAACA GTAACGGTTA AAGGTTTCTA CGGAAGAATG AAAGTTTAC	2400
CAAAAACAGG AGAAACAGAA AATATTGCAT TGTCTGTCTT AGGTTTCTTA ATGGTATTAG	2460
GATCAGCATT CATTTTTTAA AAGArAaTCT AATTAAGTAT TTATAAAAGC ATGGTCGCAA	2520
GTTTTACT CTCTACTTG AAGTtCCGA ATGAAAACCT TATAATGAGA AATGTGAAAA	2580
TGTTCACTTA TTATAAATTA AGTtTGGAGG CTTTTATTTA TGAAAATTGC AATTGCAGGG	2640
GCCGGCGCGA TGGGTTCCCG GTTCGGACTG ATGTTGAAAC AAGGTGGTAA CGATGTTTTG	2700
CTAATCGACG GCTGGCAGGA GCATATCAAT GCCATCAAAG AAAACGGATT GAAGGCTAAT	2760
TATAATGGCG AAGAAATTAC TGTAAAGTC TCAATTGTTA ATCAGAATGA GGTGCCCCACT	2820

GGTGAGCAAT TTGATTTGAT TATCTTATTC ACGAAGGCAA TGCAGTTGGA AAAAATGCTG	2880
CAGGATGTTA AACCATTAAT TGCTGATCAC ACAGAAGTTT TGTGTCTTTT AAACGGTATC	2940
GGTCATGAAG ACGTAATCGA AAAATTCGTA CCGATGGAAA AAATCTTTAT CGGCAACACC	3000
ATGTGGACTG CTGGTCTAGA AGGTCCCGGC AAGGCTAAAT TATTTGGCAG CGGGTCTGTT	3060
GAATTGCAAA ATCTAGGTAT TGGCCAAGAA GAATCAGCGA AAAAATTGGC GGAAACTTTG	3120
TCCGCGTCTG GTTTAAATGC CAAATATTCC GACAACATTC ATTATTCAAT TTATCGTAAA	3180
GCTTGTGTCA ATGGTACGAT GAATGGCTTA TGCACGATTT TAGACGTCAA TATGGCGGGG	3240
CTTGGTGCAA CCAAACCTGC TCATGACATG GTAGTGACTA TTGTTAACGA ATTTGCAGCT	3300
GTTGCTGCCA AAGAAAACGT CAACTTAGAC ATTCCCGAAG TAATCGAACA TGTAGAACT	3360
TGTTTTGACC CGACTACGAT TGGTATGCAT TTCCCATCGA TGCACCAAGA CTTGATAAAA	3420
AATAACCGTT TGACCGAGAT TGACTACATC AACGGAGCCA TTTCTCGTAA AGGTAAAAAA	3480
TATGGAGTAG TAACTCCTTA CTGTGATTTC TTAACGCAAC TGGTTCATAG TAAAGAAGAA	3540
ATTCTAGGAG CGAAATAGTA ATCATTAGAA AGAGGATGAG AAAATGGAAG AAAAATTAAC	3600
TCCAAAAATA TTTTAAATA AAGTATTAGC GGGTACCGCA ACGGGGATCA TCGTAGGGTT	3660
GATTCCTAAC GCGGTATTGG CAGCAATTTT GAAATTATTT GGTGAAAATA CCTTAGCAGT	3720
AACGATCACA CAAATGGCAG TTATTTTCCA ACTAGCTACT CCACTATTGA TTGGCGCTTT	3780
GATTGCGGTG CAGTTTGGCT TCAAACCAAT GCAAATGTTA GTAGTTGCCG GTGCAGCTTT	3840
CGTTGGCTCC GGTGTTGTGA AATTCAATCC TGACACAGCT ACTTATATCG GTGCTGGTAC	3900
CGGAGACATC ATCAATACCA TGATTACTGC ATCAATCGCT GTCGGCATGA TCTTGTTGAT	3960
CGGCGAAAAA TTTGGGTCCG TGGCAATTGT TGCAACACCA ATCGTTGTTG GTATCGGCGC	4020
TGGTCTAATC GGCTATTATT TATATCCTTA CGTAACCAAG ATTACTGCCG CTATCGGTGA	4080
CTTAATTAAT ACCTTTACTA CTTTGCAACC AATTTTGATG TCAATCTTGA TTGCATGTTC	4140
CTTTGCTTTC TTGATTATCT CGCCGATCTC TACAGTTGCC ATTGGTATGG CAATTCAATT	4200
GAACGGTGTT TCTGCTGGTG CAGCCGCTAT GGGTGTGGCC GCAACAACAG TTGTCTTGGT	4260
AGTTAACTCT TGGAAAGTCA ACAAACCAGG GGTACTTTTG GCCATCGCAT TGGGTGCTAT	4320
GAAATGATG ATGCCAACT TGTTTAGGAA ACCAATTATC TTGGTGCCAT GTCTATTCAC	4380
CGCTATCATC TCAGCAATTC CAGTAGCGTT ATTCTCAGTC TCAGGGACTC CTGCTTCTGC	4440
TGGTTTCGGT TTGGTAGGGT TGGTAGGTCC GTTGGCTTCA TTGGACGCCG GTTTGAGTAT	4500
TATTCTACTA TTAATCAGCT GGTGTGTTGT ACCAATCGTA GCTGCTTTCG TAGGTCAAAT	4560
CTTATTTGAA AAAATCTTGA AATTGTATGA TCGTAAAGAT GTTTTCGAAT TTTTAGGATA	4620
GTCAACTAGA AAAAGTGTGA GCCTTGCTGCT CACGCTTTTT TGTGCAATTG AAGTTGGTGA	4680
ATCTCAGCGA AAAGAGAAAG GTACCTTAAA TAGTGTGTTT CTGGTCAATA GCACAATGAA	4740
ATATCTAAAT ACATTATGAA CTGCTTTTCG TTTTATGAAG AGCAGTTTGT GCTATTTGAG	4800

AAATTGCTCA	ATTTAATCAT	TAAGTAGTTT	CATCAATGGT	GACTATGCTA	CTGCAATTCT	4860
TGTTGTACCA	ACCGGGAGCG	TCAATAATTT	TGTGTAAATA	ACTCGTCCTT	CTGCAAAATA	4920
ATGGGTTACT	CAGTAAACAT	TGAAGCTAAT	GTATCCGTC	CTTGTTGGAA	GCCTTTATGG	4980
CTTCTATTTA	AGAATTTTTG	ATTGTATGTA	TCAAAGATGG	ATACTAGGAA	ACGCTCTAGT	5040
GATTCTTCAT	TTTGAAACTG	CTCTTTTCTG	CGGCTGTATT	TCTTGATTG	TTTATTGAAA	5100
GATTGATTA	AATTAGTTGA	GTAAATGCTT	CGGCGGATAC	CAGGCGGAAA	CTCATAAAAA	5160
GTCAATAAAT	CTTGATTTTT	TAtyCAnkGC	TGCGTCACTT	TAGGATAAGT	TTTCTGCCAC	5220
TTCTCAATCA	TACTCCCTAA	AAAGGTATTT	GCCTCTTCCT	TTGAGCTAGC	TTGATAAACA	5280
GCCTTAAAT	CATCACAGAT	TTCTTTTCGA	TCTTTGaCAC	GTACTTTATG	AGCAATATTA	5340
CGGTAAACGT	GAACACAGCA	ATGTTGGTAT	TTAGCtTTAG	GATAAATCTG	GTGAATCGTA	5400
TCTTTCATGC	CTTTTAAGCC	ATCGGTAATA	AACAGTAAGA	CATCATGAAC	CCCTCTGGAG	5460
TAAATGTCCT	GTAACAACTC	ATTCCAAACG	TAGGTCGATT	CAGTCGGAGC	AATCGCATAA	5520
CTTAGTACTT	CTTTAGTGCC	GTCTTCTCGT	ATACCAATGG	CAATATAGAT	TGCTTCTTTG	5580
GAGACGGTTT	GCCGTTTTAA	AGAAATATAA	GTAGCATCCA	TGAAAATAGC	GACATACTTA	5640
TCATTTAGAG	TTCTAACTTT	AAATGCATTT	ACTTCTTCAG	TCTGCGATTT	CCGACATGGT	5700
AATACCTTTT	CGGAAGAGGT	GAATGACGGT	CTCCTCTAAG	GTATCATTCG	TCCGTCTATA	5760
AGCAGGAACA	GTCTGTTGCT	TGAACTCGCC	GTTGCGGTCG	CGCGGAATTC	TGTGAATGTG	5820
TTGTGGTAAC	GGTTCGTGTG	CAAAGTTTTA	AATCTACTAT	CAAATAAGGT	AGAATAATAG	5880
AAAAAGATAG	CAGGAGGAAT	GACGGTGAAT	CATTTTAAAG	GAAAGCAATT	TCAGTAGGAT	5940
GTGATTATTG	TAGCCGTGGG	CTACTATCTT	CGTTATAACC	TTAGCTATCG	TGAAGTTCAA	6000
GAAATCTTAT	ATGATCGTGG	CATTAACGTT	TCTCATACGA	CGATTTATCG	TTGGGTGCAA	6060
GAATATGGCA	AACTACTCTA	TCAAATTTGG	AAAAAGAAAA	ATAAAmATC	CTTTTATTCA	6120
TGGAATAATGG	ATGAAACGTA	CATCAAAATT	AAAGGaAAAT	GGCATTATTT	GTATCGAGCC	6180
ATCGATGCAG	ATGGTTTaAC	CTTGGaTATT	TGGTTACGTA	AAAAACGGGA	CACACAAGCA	6240
GCCTATGCTT	TTCTTAAGCG	GTTAGTGAAG	CAGTTTGATG	AACCGAAGtT	GtAGTCACAG	6300
ATAAGCCCC	CTCTATTACA	AGTGCCTTTA	AGAACTAAA	AGaATACGGC	TTTATCAAG	6360
GGACAGAACA	TCGTACCATT	AAATACCTGA	ATAATTTGAT	TGAACAAGAC	CATCGTCCAG	6420
TAAAGAGACG	CAATAAATTC	TATCGAAGTT	TACGCACTGC	CTCTACCACG	ATTAAAGGCA	6480
TGGAAGCCAT	TCGAGGATTA	TATAAGAAAA	CCCGAAAAGA	AGGCACTCTC	TTCGGGTTTT	6540
CGGTCTGTAC	TGAAATCAAG	GTATTATTGG	GAATCCCAGC	TTAAATCATA	GATACCGTAA	6600
GGgATTTTAT	TCTTTATTTA	AAACTTTGCA	ACAGAACCAA	TAATTTGATC	TACTCATTTT	6660
TAATTCTTGG	GAGAATTTTT	CTTTTGATTT	GGTACCTCTC	ATCTTTTTTA	AGCTTTCTCT	6720
AAATGCTAGT	TCCAAGTTTC	TCCCTCCTGT	TTATGTAAAT	ATTTTGCACG	TTTTCCATTT	6780

AATTATATCG TTACACCTTG TAAATCACAA GTGATTAATC ACAAATCACT TGTGATTAGT	6840
GATTGTTATT .GGTGTTCCT AGAACGATTC TAAGCACATG TTTAATTTTA AAGTGGGCAA	6900
TTATCTTTAA AATGTCCTAC AATGATTCAA TAAAAGCATG GCTGCTAGTG AAACCTGCGA	6960
CCATGCTTTT AAAAATACTT AATGGGGTCC CGAGCGCTTT AGCTCCTTGG AAGCTGTCAG	7020
TAGTATACCy AATAATTTAT CTACATtCCC yTTAGTAACG TGkAACTTTC CAAATTTACa	7080
AAAGCGACTC AtAGmmTtAT TTCCTCCCGT TAAATAATAG ATAACtATTA AAAATAGACA	7140
ATACTTGCTC ATAAGTAACG GTACTTAAAT TGTyyACTTT GGCGTGTTC ATTGCTTGAT	7200
GAAACTGATT TTTAGTAAAC AGTTGACGAT ATTCTCGATT GACCCATTTT GAAACAAAGT	7260
ACGTATATAG CTTCCAATAT TTATCTGGAA CATCTGTGGT ATGGCGGGTA AGTTTTATTA	7320
AGACACTGTT TACTTTTGGT TTAGGATGAA AGCATTCGCG TGGCAGCTTA AGCAATTGCT	7380
GAATCGAGAC TTGAGTGTGC AAGAGCAACC CTAGTGTTCT GTGAATATCC AAGGTACGCT	7440
TGTAGAATCC TTCTTCAACA ATCAGATAGA TGTGACGCG ATGGCTTTCA AAAACCACTT	7500
TTTTAATAAT TTGTGTGCTT AAATGGTAAG GAATATTCCC AACAATTTTA TACCTCTGTT	7560
TGTTAGGGAA TTGAACTGT AGAATATCTT GGTGAATTAA AGTGACACGA ATGTTCACTT	7620
TTAATTTTTT TGACGATAAG TTGAATAGAT GACTGTCTAA tTCAATAGaC GTTACCTGTT	7680
TACTTATTTT AGCCAGTTTC GTCGTTAAAT GCCCTTACC TGTTCCAATT TCGTAAACGG	7740
TAwCGGTTTC TTTTAAATTC AATTGTTTTA TTATTTGGTT GAGTACTTTT TCACTCGTTA	7800
AAAAGTTTTG AGAATATTTT ATATTTTGT TCcATGTAAT CACTCCTGAA GTGATTACAT	7860
CTATAATAA ATACAGAAGT TAAACGATTT GTTTGTAATT TTAGTTATCT GTTTAAAAAG	7920
TCATAAGATT AGTCACTGGT AGGAATTAAT CTAACGTATT TATTTATCTG CGTAATCACT	7980
GTTTTTAGGG TTCTGTTGCA AAGTTTTAAA TAAAGAATAA AATCCCTTAC GGTATCTATG	8040
ATTTAAGCTG GGATTCCCAA TAATACCTTG ATTTCACTAC AGACCGAAAA CCCGAAGAGA	8100
GTGCCTTCTT TTCGGGTTTT CTTATATAAT CCTCGAATGG CTTCCATGCC TTTAATCGTG	8160
GTAGAGGCAG TGCGTAACT TCGATAGAAT TTATTGCGTC TCTTTACTGG ACGATGGTCT	8220
TGTTCAATCA AATTATTCAG GTATTTAATG GTACGATGTT CTGTCCCTTG ATAAAAGCCG	8280
TATTCTTTTA GTTTCTTAAA GGCACCTGTA ATAGAGGGGG CTTTATCTGT GACTACAACn	8340
TTCGGTTTAT CAAACTGCTT CACTAACCGC TTAAGAAAAG CATAGGCTGC TTGTGTGTCC	8400
CGTTTTTTTAC GTAACCAAAT ATCCAAGGTT AAACCATCTG CATCGATGGC TCGATACAAA	8460
TAATGCCATT TTCCTTTAAT TTTGATGTAC GTTTCATCCA TTTTCCATGA ATAAAAGGAT	8520
TkTTTATTTT TCTTTTCCA AATTTGATAG AGTAGTTTGC CATATTCTTG CACCCAACGA	8580
TAAATCGTCG TATGAGAAAC GTTAATGCCA CGATCATATA AGATTTCTTG AACTTCACGA	8640
TAGCTAAGGT TATAACGAAG ATAGTAGCCC ACGGCTACAA TAATCACATC CTGCTGAAAT	8700
TGCTTTCCTT TAAAATGATT CATCGTCATT CCTCCTGCTA TCTTTTTCTA TTATTCTACC	8760

TTATTTGATA GTAGATTTAA AACTTTGCaA GAGAACCGTT CATGGGTATA GCATGAATTG	8820
TTTAACTGTC CTCTAGGATT CCCGTCCTAT GGCAGTCGGT TCTTATTTTG TTAATTGTCT	8880
TGTAAGTTGC CTTTATATCG CACTAAGTTT AAAAAAGTAA AAAGATTAT CTCTCAGATA	8940
ATTGCTATAT CAAGGTAAC ACGAGATTCT AAATAATTA TCCTACAAGT GTGACAACAA	9000
ATAAAGTTAT CAATGGAATA ATTTTTCAT TAGCTCGTTT GCTCATTCTA ATAGCTCTCC	9060
TCTTATTTAA ACTTAACTTA CAAATTAATA ATAGCATAGA TTAAATATA AATAAACAA	9120
AATGCAAAAA TTTGCGCTTT TTCATCAAAA AAATTTATTA CATAATATA AAACCAAAAA	9180
ATCATTCCAC GAATGCACTT TTTTGC GCCT CGGAAGGAGT ATTTCAAAAT GTTCTTTAC	9240
GAATATTCA AAGAACAGCG AAAACTCTAT AATTTTACTC AAGAAGAATT TTATGAAGGA	9300
ATATTCAAAA AAAGAGCTGC CTCTCTTTT GAGGTACACA ACACGCATGA TCTCAAAGTT	9360
AAAGATTTAC CTGTTTTATC AGATAGAAGC ATGATGAGTA TCTTAGAAAT AATTCATTAT	9420
GCTAAAGAAG AATTTATTTT TCCtTACGAC GAAGACTTAA ATTCATTGTT TGATATATTC	9480
CAGAATAATA CCGAAAAAGA AAATAAAGAC TACATTTACA AATTATACAA AAAGTCTATA	9540
GAGCTAAAG AACATTCTAT AATATATTGG AATCTTTACT TAATAATAAA AATCCAATGT	9600
TCTGAATACG ATTCAAGAAT TGTCCCTACA GACTCACAAG ATTTATCTGA GTTAAAAAA	9660
ATGaTTTTAT CAAaACaAAA GTTTACTCTT TACGACTATA AAATTGTTAC TAATCTATCT	9720
CTTGATTTTT CTTACAAAGA ATTGCAACC	9749

## (2) INFORMATION FOR SEQ ID NO: 155:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9021 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

AAAATCAGAT ATCTTATAAn nGnAATCAAG CACGTATTCG CGATAAGGGT AAAAAATTAC	60
GTAAGAATCG GCACAGCGCT TCTTTAAAAA GTAAATTAAA CTTTGACAAC TTTTTCGAAT	120
AATGGTACTG TTAAGAAGTA AATTTGTGAA TAGAGGAGGT CGAAGAACAT GGCAACAAAG	180
aAAGCAGCTC TTGCTTGTTT CGTTTGTGGC TCTCGCAACT ATTCTAAATC CGTTAGTGAA	240
GGTAAACGTG GCGAACGCTT AGAAATCAAT AAATTCTGTA AATATTGTAA TCAATACACG	300
TTACATAAAG AAACGAAATA GATTTAGAGG AGGTATCAA CATGAAATTT TTCCGCAGCG	360
TTGCTGATGA GATGAAGCAA GTAACCTGGC CAACAAAAAA ACAATTGCGT AAAGATACAT	420
TAGTTGTAAT TGAACATCT ATTTTATTTG CAGCATTGTT CTTTATCATG GaTACTGTTA	480
TCCAAACGGC ATTTGGcTGG ATTTTAAAT AATTAATTCT AGTATTTCCC GACGAGGGGT	540
GCTATAATAG TATTCGAGAA AAAGCTTCGG GAAACTGAGC TTTTTTTATT TGCTCGAAAA	600
AAATAAACTA AAACATTTAA AGGAGCCGAA TCAGTAATGG AAACTTTGA AAGAAATTGG	660



TAcGTATTGC ACACCTATTC AGGCTATGAA AATAAAGTAA AwGCAAACAT TGAATyACGT	720
GCACAAAGCA TGGGGATGGG CGATTATATT TTTCGTGTCG TTGTACCAGA AGAAACAGAA	780
AAAGAAGTCA AAAATGGTAA AGAAAAAGAA ATCGTTCATA AAACTTTCCC tGGTTATGTG	840
TTAGTGGAAA TGATTATGAC CGATGATTCT TGGTATATTG TTCGTAACAC GCCAGGCGTA	900
ACTGGTTTTG TAGGCTCACA TGGTGCTGGT AGTAAACCCG CTCCTTTATT ACAAGAAGAA	960
ATTAACCACA TTTTACGTTC AATTGGCATG AGCACCCGTC AATCAGACTT AGAAGTAGCG	1020
CTAGGCGATA CAGTGAAAAT TATCGAAGGT GCCTTTTCTG GTCTTGAAGG TGTTGTGACA	1080
GAAATCGATG AAGAAAAACA AAAATTAAAA GTAAACATCG ATATGTTTGG TCGTGAAACA	1140
AGTACAGAAT TAGACTTTGA ACAAGTCGAT AACATTGATT AATAACAAG CTGACAACT	1200
ATTGTTcAGC TGATTGATAC AGATTGTAGG GTCTGAAAA GCAGTCAACT GACTTCTTTT	1260
CCAAGCCCTT TTTTAGTTCA TCGCAGAAAG GatGrAAAA AATGaACATG CCCAAAAATA	1320
TChGTtATTT TTCTTTGCTA ATGGGTCTtG tTCTATTATT AAGTGCTTGc CAAATTGGGG	1380
CaACTACGAA GGATGACAAC CAAGCCGCCa CAAAAGAAGC AACTGTTGAG TTAAACCGCA	1440
CAACAACACC AACGCTTTTT TTTCaTGGTT ACGCAGGAAC TAAAAATTCG TTTGGCTCGT	1500
TACTGCATCG CTTGGAGAAA CAAGGTGCCA CAACTCAAGA ATTAGTGCTA CTCGTTAAAC	1560
CTGATGGGAC CGTGGTTAAA GAGCGAGGAG CTTTAAGTGG CAAAGCGACG AATCCCAGTG	1620
TTCAAGTTCT ATTTGAAGAT AATAAAAACA ATGAATGGAA TCAAACAGAA TGGATAAAAA	1680
ACACATTACT CTATTTACAA AAAAATTATC AAGTGAACAA AGCCAATATT GTCGGGCACT	1740
CTATGGGTGG TGTTAGTGGT TTACGTTATT TAGGAACCTA TGGGCAAGAT ACATCGTTAC	1800
CTAAAATTGA AAAATTCGTC AGCATTGGAG CACCTTTCAA TGATTTTATT GATACGAGTC	1860
AACAGCAAAC CATCGAAACG GAACTAGAAA ACGGCCCCAC AGAAAAAGT AGCCGCTATT	1920
TGGATTATCA AGAGATGATT AATGTTGTTc CAGAAAACT GCCCATTTTA TTAATTGGTG	1980
GTCAATTAAG TCCAACAGAT TTAAGTGATG GAACGGTGCC GTTATCTAGT GCCTTAGCAG	2040
TCAACGCCTT GCTAAGACAG CGAGGAACTC AAGTCACTAG CCAGATTATT AAAGGAGAAA	2100
ATGCACAACA TAGTCAATTA CATGAAAATC CTGAAGTAGA TCAATTGCTA ATCGAATTTc	2160
TATGGCCGAG TAAAAAaTAG CAaAAAGGCA CATGTTCCAT AAAAAATGGA ACATGTGCCy	2220
TTTTCTATAC TTTGGCTAGC GAAATAGCTT TTTTCAGCTA TAATAAATAA AATTATTTAT	2280
TTTTATGAGA TATTGGAATG GAATCCTCAT GATTTAAATA AGCACAAATC AAGGCTTCAA	2340
AGGTGATTAA TGTTTGGCCA AAAAACATTT GTGGCTGAGC GACACTTGTT GAAAACGGGC	2400
TGTAAGAGTC AGTTGAAAAA ACGAGATCTG CCAGTCGACC AAGAGTTGAA TTTTGGCGC	2460
CAACAAAAGA AATGATAGGA TGTTTTTTCT CTTTCGCAAG CAGAGCTGTT TCTTTAAGT	2520
ATTTTTCTTC ACCAGAATGT GAAAGGATGA TAAAAAGAAA ATTCTGATTA TTTTGAATAT	2580
TTATCGAATG AGTATAGGCT GTTGTGATGC TGGGGATGCC ATGAAAATTG AAAACTTCAC	2640

TAATGTAAGC CGCCAAATGT CGTGAAAAGT CATGACCTAA AATAACAAAT AAATGACTCT	2700
TGTGTTTGGC AAGCAGTCGA CAAAACTCGT TGGTTGTTTC AAATGAAGGT GCTGTTTCAA	2760
AAGGGATTGG CTGAGAAAAA TGGGCTTCTT TTAATTTGTA AATAAGTTCA TTATAACCTG	2820
ATAAATTTAA TTTTTTGGCT AAACGAATGA TGGAAC TAGG CGAAGAAAAG CAGTGTTTTG	2880
CAACTTCCCG AATACTCAAA TCGGCCGCAT TCGTACTTTG TTGATCCAAA TAATACAAAA	2940
TCTGaCTTTC TGTTTTCGTT AAATGaTACG TTTCTTTTAA GTAATCTAAG TTCATGAGCA	3000
AACCTCCTTT ATCAGATTAT AAATAATTTT GTGGGAAAAA GGTATCAAAA AATAAAAAAA	3060
TGAGGTGTTT TTATGAAAAG GGCAACAAAG CAAAGGCTGT CTTTGGCAGC AATCATGGTT	3120
CTACTTCTCT CGGGCTGTGG AAGTGTGGG AAAGAAACCA AAAAGCAAGA ACAACAGGTA	3180
TTACGGGTCG GGATTGATTC GGAATTATCA ACGGCAGACG TGTCGTTGGC AATGGATAAT	3240
ACCGCAGCAG ATGTAATGAG CCAAGTAGGG GAGGGACTTT TCTCCTTTGA CGAAAAAGGA	3300
GAAGCGAAAC CAGCATTGGC AACTGAAAAA GTACAGCCCT CCAATGATGG TTTAAGCTAT	3360
ACTTTTACGA TTCGAAAAGA TGCAAAATGG AGTAACGGCG AGCCAATCAC AGCAAATGAT	3420
TTTGAATACT CTTGGAAGCG CACAGTGGAC CAAAAACAG CTTCCCCGCA AGCGTATTAC	3480
TTTGAAGGGT TAAAAATTA TCGTGCTATT GTTGACGGTA GCAAATCTAA AGAAGAGTTA	3540
GGGGTAACAG CCATTGATGA CCATACCTTG GAAGTAGAGC TAAGCTATCC TATGAGTTAT	3600
TTTCAACAAT TATTGGCGGT ACCAGCTTTT TATCCTTTAA ATGAAGCATT TGTCGAAAAA	3660
ACGGGCAAAA ACTATGGTAC ATCAGCTGAG TCAACACTTT ACAATGGCGC CTTACATTA	3720
GAAGGTGGG ATGGCACGAA TAATACTTGG TCCTATGTGA AGAATAAAAA TTATTGGGAT	3780
CAAGCGAATG TTTGCTAGA TAAGGTGGAT GTCCAAGTAG TTAAAGAAGT CAATACTGGG	3840
AAAAATCTTT TCGAAGGGAA AGAATTAGAT GTTGTA AAAA TTTCTGGAGA AATTGTTGCA	3900
CAAGAACAAG GCAATGCAGC TTTGAAAATT CGTGAAATTC CTGGAACGTA TTATATCCAA	3960
TTAAATACGC AAAAAGATCT TTTGGCAAAT AAGAATGCAC GTCGAGCAAT AGCATTATCA	4020
TTGAATTCTG AGCGTTTAGC TAAAAATGTT TTAAATGATG GCTCAAAAAA AGCACTTGGC	4080
TTCGTGCCAA CAGGTTTCAC TAATCAAGAA ACGCAAAAAG ATTTTGCAGA GGAATTAGGA	4140
GATTTAAATC CTAGTGAACC AGAAAAAGCG AAAGAGTTAT GGCAAACGGC TAAAAAGAA	4200
TTAGGAATTG AAAAAGCGGA GcTAACGATT TTAAGTTCGG ATACAGAAAA TGCTAAAAAA	4260
ATCAGTGAGT ATGTTCAaGG AGCTTTAGCA GATAATTTAG AAAATTTAAC AGTCAATGTT	4320
TCACCAGTTC CTTTAAATAA TCGTTTAGAA AAAAGTCGCA GCGGAGATTT CGACATTGTG	4380
GTTGGTGGCT GGACGCCaGT ATATGCTGAT CCAATCGATT TCTTAACTT ACTGCAATCA	4440
AAAAATTCCA ATAATTTTGG TAAATGGTCT AATAAGACCT TTGATCAGTT GCTTCAAGAA	4500
GCAAACGTAA CTTATGCAAA TAAATATGAA GAACGTTGGA AAACATTACA AAAAGCGGAT	4560
CAATTGGTTG CGGAAGAAGC CCCCCTAGTT CCTCTTTATC AATTAACAGA AGCACGCTTA	4620

GTGGCCGATT CTGTCCAAAA TTTAGTCTAT GGTCCATTAG GTTCAGGCTA TTACAAATCA	4680
GTCTCTATCG GCGACAAGTA AAGGTTGGGA AAAAATGAAA AAAAGTCAAG CTGTTCAATT	4740
AATTCAGGAA CTTTCAAATG CTAATGGTGT TTCAGGCTTT GAAACAGAAG TCGTTAGAAT	4800
CCTTCAACAC GCAACTGCGG ATTTTACTAT TCAACGGCTA GATTCCATTA AAAATTTATA	4860
TCTTGAGAAG AAAAATAATC TAAGTGAAGG TCCAGTAGTG CTTTTTGATG CTCATAGTGA	4920
CGAAGTGGGT TTTATGATTC AAGCAATTAA AGAAAACGGC TTGTTGCGTT TTTTCCGTT	4980
AGGTGGCTGG GTCCCTAATA CGATTCTGC AAAAAAGTT CGGATTGCGTA ACCGGGAAGG	5040
TACCTATCTG CCAGGCGTAG TTACTAGTCG GCCCCCTCAT TTTATGACAC CAGAGGAACG	5100
TCAGCGACCG TTAACCATTG CTGATTTGAC GATCGACATT GGTGCTACCT CAAAAGAAGA	5160
AGTAATTGAA ACTTACAAAA TCGATCTTGG TGCACCGGTC ATTCCAGATG TGACGTGCTG	5220
TTACAACGAA CAGACAGATC TTTTTTTAGG AAAAGCTTTT GATTGTCGCA TTGGCTGTGC	5280
TTGCTTAGTG GATGTCATGG AGGAGTTGAA AGAAGAGACG CTTCCATTCA AGTTAGTTGC	5340
GACAGTAACT GCACAAGAAG AAGTTGGCGA AAGAGGAGCA CTTATTGCCG CAAAACAAGT	5400
CAATCCTGAT TTAGCTATTG TTTTCGAAGG CTGCCCGGCT GATGATACGG CAGAAACGCC	5460
AGAGATGATT CAATCAGCAA TGGGAAAAGG ACCGATGTTA CGTTATTTTG ATGTTTCTAT	5520
GATTACGAAT CCGGAATTC AAGAGTATGC ACTAGAGATT GCTAAAATAC ACAAATTC	5580
TGTTCAAGTT TCTGTTGAA GTGGTGGTGG AACCAATGGC ATGGCGATTA CACAAGTTCA	5640
AGGAGCCCCG ACGATTGTTG TAGGAATTCC TGTCCGTAC GCTCACACAC CCCATTGCTA	5700
TGTAGATTTT CAAGATTACC AAGCGCGGAA AGAATTAGTC ATTCAATTAA TCAAAAATTT	5760
AGATGCTGAC AAAATTCAAG CACTGGTTCA GCCATTGTCA AAGGAGTGA ACAAATGAAG	5820
ATTTATGTAT CATGCGATAT TGAAGGATTA GCAGGTATTG CGACATTTGA TATGGAAAAA	5880
GAAGACACCG TTTTATTCCG AGAATTATAT CACCAACATG TGGCTTGGTT AATTGAAGGG	5940
ATTCAAAAGA GTGCCAAAAA TGAACAAATT ACAGAAATTA CTATTGCCGA TTCGCATAGC	6000
CGTGGATTGA ATTTAGCTTA TGCGCGCTTG GCAGAAATGG ATGAACGAAT TTCCTTAGTG	6060
AGCGGCTTTC CTCGCATGGA TTATATGATG AGTGGCTTGG ATAGTAGCTA TGATGTGGTC	6120
TTTTTCTTAG GTTATCATGC CGGTATTGGT AAACAAAAG GGAATATGGA TCATGGGTAT	6180
AGCGCTAGTG TTGCTTATGA TTTAAAAATA AACGACTTAG CGATGAATGA AACGACAATT	6240
AATGCAGcTT ACGCTAGCGA ATTAGGTGTC CCTGTTGGTT TAATTATCGG TGAATCAGGC	6300
TTAGAGGAAC AACTTTTTCA AGAAAAATG ATGCCGAAG TTCCCTTTGT TTCCACCAAA	6360
GAATCTCTGG GACGTTATGC AATCAAAAAT AGACCCATGC AACAAAGTTCG GGAAGCTATA	6420
GTAGCAACTA CAAGTCAGGT GTTAACTTCC TTTGCGTTGT CTGAATTACC CCGATATGCC	6480
TTGCAAACAC CTGCAACTGt AAAATTGCAA TGTGTGACGA CAGCACAAGC AGATCGTATT	6540
GAAATGTTAC CAATGATTAA ACGCATTGAT GGTAGAACTG TTTCATTTGT AGGTGAAACG	6600

ATGAAAGATG TCATGAATGG AATTGTCGCT GTGGTTGGTT TAGGCGGAAC AAGTTATTAA	6660
AGTAAGAAGC TTACGGAAAA AACGAATGGA CATTTTTTTC TGTAAGCTTC TTGCTATAGT	6720
TATTTCTATT TTTGTTTCTC TAGAATTTCT TTGCTTAAAC GTTGTGCAGT TGGTGTCAAC	6780
GCAAGACcGC CTTCAGCGGT TTCCTTGAAA ATTTGTGGCA TTTGGCGTCC CACTTGGTAC	6840
ATTGCGTGGA TGACTTCATC TGGCGGAATG ACACCTCGAA TCCCCGCTAA AGCCATGTCA	6900
GCTGAGATGA AGGCTTGAGA AGAACCTAAT GCGTTTCGCT TCACGCAAGG GACTTCAACA	6960
AGTCCCGCAA CAGGATCACA AATGAGGCC ATCATATTTT TTATCGTAAT GCCCACAGCT	7020
TGAGCAGATT GATCAGgCGT ACCGCCAGAA GCCGCACTA AAGCGGCACT TGCCATTGCA	7080
CTAGCAGAAC CGACTTCAGC TTGACAACCA CTTCGGCACC ACTGATTGAG GCATTATTGG	7140
CAATCACTAA GCCAAAAGCC CCAGCTGTAA ATAGAAAATC CAATTGTTGT TGATGCGTTA	7200
ACTTTAAGCG ATCAATAACT GCCAACAAGA CGCCAGAAAC TACTCCGGCA CTGCCTGCGG	7260
TTGGTGTAGC ACAAATCAAG CCCATTTTTG CATTAACTC ATTAACAGCA ATCGCATTTT	7320
GCACGGCTTG TAAATGGTC TCACCGCTTA AAAAGTTGCC TGAATGTATG TAATCATTTA	7380
AGCGGGTTGC ATCGCCACCA GTTAGCCCTG TTAGTGAGGT AACACCAGCG ATTCCTTCAG	7440
CGACGGATTG TTTCATAACG GCTAGATTTT TTTCCATTGT TTCAATAATT CGTTCACGAG	7500
AATACCCGCC ATGAGCCATT TCTGTTGCAA TCATTAATTC GCTCACAGAA GGGTAATCTT	7560
TCGCTTGAGC TACCAAATCT TCAATTGAAT AAAACATCGT ATTACTCCTT TACGTTTATT	7620
TAGTCAAAAT AGTTAACGCT ATATATATGA GGGATTTTCAG CTAATTGCAT GACAATGTCA	7680
CCGACTTCGG CTTGATCAAC TTCAATAATC ATAATCGCTT TTTCCCTTT AGATTCTCGG	7740
GTAACGGTCA TTGTACTAAT ATTGATATCG CTAGCTGAAA GGATGTTGGT CACCTTAGCA	7800
ATCATGCCAG GAACATCTTG ATGAACAACG ATAAATGTTG GCGTGCCCAT ACTTAAAGAC	7860
AATTTGAAGC CATTTAATTC AGAAATTTGA ATGTTTCCAC CACCAATTGA AATTCCTGTC	7920
ACAGATAGTT TTCGGTCGCC AGAAGAAACC ACAATTTTAA CGGAATTAGG GTGGTCTGCT	7980
TTTTCACTTT TAGGAATAAA ACAAACCTCC ATTCCTTGTT CATAGGCAAT TTCCAAAGAA	8040
TTGGATAGTT GCTCATCATC TGGTTCCATG CCTAAAAGGC CGCCAACTAA GGCGATATCT	8100
GTTCCGTGAC CACGATAAGT TTTGGCAAAA GACTCATATA AATAAATGTC AACGCTATCA	8160
GGTTGCTGTC CAAAATACT GCGAACCCT TTTCCAATTC TTGCCGCTCC AGCTGTATGA	8220
GAAGTCTTG GACCAATCAT AACAGGTCCG ATAATATCAA AAACACTATT AAAGCGTAAC	8280
TCTTCCATT CAGCACCTTC TTCATTAGAA TATGTTTTCA TTGGAAGTCT AACACAAAAA	8340
ATGGCGGGGA ACTAGGCAAT CGCTTGTGTT TTAGCGAATA ATTTAAATTT TTATCAGAAT	8400
GAAAGCAAGG GGCCTTCGTT GACAAATCAG TGAAGAGATG CTAAATTCTA GATAATAAAA	8460
GACGAAAAGA GGGATTCTGT CGTTGAAAAT TTAAGTTAAT TTTTATAGAGA AACTGTATAA	8520
CTTTGCCTGA ATATAACGGG TGGAGTGTGC GGATTTTCGG AAATTAGCTT AAAGAATTCA	8580

885

ACCACGGGTC TTTTGTATA CCTTGGATTG ATGCCGCTGA TTTCCGTAAA GCTGCATCAA	8640
TTTTGCTTTA TGGAGGTATA TTTATGTCGA AAATTGAACT AAAACAACTA TCTTTTGCCT	8700
ATGaTAATCA AGAAGCGTTG CTTTTTgATC AGGCAAATAT CACGaTGGAT ACCAATTGGA	8760
AATTAGGaTk GATTGGCCGC AATGGCCGTG GGAAAACAAC CTTATTAAAGA TgTTACAAAA	8820
GCAGTTGGaT TACCAAGGAG AGATTcYtCa TCaAGTCGAT TTCGTCTATT TTCCACAAAC	8880
AGTTGCAGAA GAACAACAGC TCACCTATTA TGTCTTACAA GAGGTGACTT CTTTGAACA	8940
GTGGnAATTA GAACGAGAAT TAACGCTTTT AAACGTTGAT CCTGAnGTTT ATGGCGTCCC	9000
TTTCTTCTTn ATCAGGCGGC G	9021

(2) INFORMATION FOR SEQ ID NO: 156:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6491 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

ACAATAAACG GGAGGGATCG TTTATGAATA TTGAAAAAT GACAACCACG CTACAAsAGG	60
CGATTGCTGA AGCTCAAAAA GTTGCGGTGA CACGTCAACA TCAAGAAATT GATATCGCAC	120
ATTTATGGAA AATTTTCTT CAACCAAAACC ATTTTGGACG TAATTTTAC ACAGATGCTG	180
GTTTAGATGT TGACGCATTT GAACGAGAAG TAGATAACGC GTTAGATGAA TATCCAAGTG	240
TGGCAGGCGG TAATGTGCAA TATGGTCAAA ATTTAAGTCA AAATTTGTTT CATTATTGC	300
AAGAAGCTGa TTCACTTCGA GAAGAtTCCA AGATGaATTT TTATCaaCCG rAATTGtACT	360
TTTAGCTTTA ATGrAATTGa AAAATTATCG TTTaACmAAA TATTTAATGc mACmAGGcAT	420
TACGGaAAAA GAGTTwAGrA AAAATATTGA AGaGATGAGA GGAGGAGATC GTGTGACTTC	480
TCAAAACCAG GAAGAACAAT ATAAAGCACT AGAAAAATAC GGTGTTGACT TAGTACAACA	540
AGTAAAGCA GGAAACAAG ATCCAATCAT TGGTCGTGAC GAAGAAATTC GTGACGTTAT	600
TCGGATTTTA TCAAGAAAAA CTAAAAATAA TCCAGTCTTA ATTGGTGAAC CAGGTGTTGG	660
TAAACAGCG ATTGTTGAAG GATTAGCACA ACGAATCGTT CGTAAAGATG TTCCCGAAAA	720
CTTAAAGAT AAAACCATTT TTTCTTTAGA TATGGGCGCC TTAATTGCGG GAGCAAAATT	780
CCGTGGCGAA TTTGaAGAAC GGTAAAAGC TGTTTTAAAA GAAGTGAAAA AAAGTGATGG	840
CAAAATCATT TTATTCATTG ATGAAATTCA TAATATCGTC GGAGCTGGTA AAACAGAAGG	900
CAGTATGGAC GCCGGAAATT TATTA AAACC GATGCTAGCA CGTGGTGAAT TGCATTTAAT	960
CGGTGCCACA ACGCTTGATG AATATCGCCA ATATATGGAA AAAGATAAAG CCTTAGAACG	1020
TCGTTTCCAA AAAGTATTAG TCAAAGAACC AACAGTAGAA GATACAATTT CTATTTTGCG	1080
TGGCTTAAAA GAACGATTTG AAATTCACCA CGGCGTGAAT ATTCATGACA ATGCTTTAGT	1140

TGCTGCCGCT	ACTTTATCCG	ATCGCTATAT	TACTGATCGT	TTCTTACCAG	ATAAAGCGAT	1200
TGACTTAGTC	GATGAAGCAA	GCGCAACGAT	TCGAGTGGAA	ATGAATTCGA	TGCCAACAGA	1260
GCTCGATCAA	GTAACACGTC	GGTTAATGCM	ACyAGaAAtT	GAAGAAGCGG	CCTTGAAAAA	1320
AGAATCAGAT	GATGCAAGTA	AAAAACGGTT	AGCCAACTTG	CAAGAAGAAT	TAGCAGATTT	1380
GCGTGAAGAA	GCCAACTCAA	TGAAAATGCA	ATGGGAAACG	GAAAAAGAAG	AAGTTAATGC	1440
GGTTTCCAAT	AAACGTGCAG	AGATTGATAA	AGCAAAACAC	GAATTAGAAG	ATGCTGAAAA	1500
TAATTATGAT	TTAGAACGAG	CTGCTGTTTT	GCGTCATGGA	ACAATTCCGC	AATTAGAACA	1560
TGAATTGAAA	GAATTGGAAG	AAAAGAACGC	TAAAGATAAC	GTCAAAATGG	TACAAGAATC	1620
GGTTACTGAA	AACGAAATTG	CGCAAGTGGT	CGGTCGTTTA	ACTGGCATTG	CTGTGACAAA	1680
ATTAGTTGAA	GGCGAACGAG	AAAAATTAAT	GAACTGAAT	GAAACATTAC	ACAAACGTGT	1740
GATTGGTCAA	GATGAAGCCG	TCGATGCTGT	TAGTGATGCG	GTGATTCGTT	CAAGAGCAGG	1800
CTTACAAGAT	CCAAATCGcC	CACTCGGTTT	GTTCCTTTTT	CTAGGACCAA	CTGGTGTTGG	1860
TAAAACAGaA	CTTGCTAAAG	CTTTAGCTGA	AGATTTGTTT	GATTCTGAAG	ATCATATGGT	1920
ACGGATTGAC	ATGAGTGAAT	ACATGGAAAA	ACATGCCGTG	TCTCGTTTGG	TTGGTGCCCC	1980
TCCaGGCTAT	GkTGgTTwTG	AAGAAGGTGG	CCaGTTAACG	GAAGCTGtTC	gTCGrAACCC	2040
tTATACAATT	GTCtTATTAG	ACGAAATTGa	AAAAGCGCAC	CCAGATGTCT	TTAATATCtT	2100
ATTACAAGTT	TTAGATGATG	GTCGTTTGAC	AGATTCTAAA	GGTCGGGTCG	TTGATTTTAA	2160
AAATACAGTT	TTAATTATGA	CAAGTAACAT	CGGCTCTCaG	CTTTTATTGG	AAGGTGTGAC	2220
ACCTGAAGGA	ACAATTCAG	AAGAAGTTGA	AAATCAAGTC	ATGAATATTC	TAAAAGGACA	2280
CTTTAAACCA	GAGTTCTTAA	ACAGAATTGA	TGATACCATT	TTATTTACAC	CATTAAGCTT	2340
GGaTAATGTG	AAAGGAATCa	TTGGTAAAAT	GACCGCACAA	CTTGCTCATC	GTTTAGAGCA	2400
ACAAGAAATT	GTGTTGGAAA	TCACTGATGA	AGCTAAAACA	TGGATTGCTG	AAAATGGGTA	2460
TGAGCCAGCT	TATGGAGCTC	GTCCGTTAAA	ACGTTTCATT	ACTCGCGAAG	TCGAAACACC	2520
ATTAGCAAAA	GAAATTGTTT	CTGGACGAGT	AATGCCAAAA	ACAAAAGTCA	CGATTAGTCT	2580
ATTAGACAAT	CAATTAGTTT	TTGAAAATGA	ACCGATAGAA	GAAGTGTAAG	AAAGAACGAG	2640
CGAATTCGCT	CGTTCTTTTT	TTATTATGAC	AATCGTTGCT	TAATTTCTTC	CAACAAATCC	2700
TCTTCTGAAG	AAACAACCCG	ACCATTCAAT	TTAATCAAAC	CAACTGTGTA	AAGGTTTAAA	2760
TAATGGAATT	GATTTTCAGC	GATTTTCGTT	AAAGCATCAA	GCTTTTGGTG	ATTATCAGCA	2820
CCTTGTTGGC	GGCTGTCTGT	GTAAAGGGCA	ACCACCGGGA	TTCCTTTGGC	ATAAGCAACA	2880
CCAATTTCTG	AAGcTACGCC	CGCATCAATG	GTTGGTCCGT	CTAATAAAGC	AACCAATAAG	2940
TCGCTAGCAA	GGACGTTTTT	TGTATCTGCT	AAGGCAATCA	TTTTGCTATC	GGCATAAGCA	3000
GATTTGTCGT	TAATCGCCGC	ATTTTCTTGT	GGTAAATAAA	GATCGATTGT	TTTATCCAAT	3060
TGTCGAATTT	GTTGACTAA	ATAGGCATTA	TAACGTAAGT	CAGCTTGCGA	AAATAAAGGT	3120

CCTGCAAAAT AAATTTTGT CATTGTCTG, TCTCCAATCC AAGGGTTTAG AAATAATCAC	3180
TGAAATAAGC TTCTTGGTCA GGAAAAATCC GTTCCAAAGA ATTTAAAGTT TCTTTATCTG	3240
TATCAATACG TTCAATTCGA TGTAAGTACG AAGGGCGTCG GTAATTCATG ACTAAACAAG	3300
TTAGTGTTTG AATATCTAAG TGAACGGCGG TGCCTAGAGG TTCATCCGTA ATGGTCACTT	3360
GATCATTTTC ATCCCAAATT AAGCCAAAAA TACCATTATT CCACTCAGCG ACAGGATCTT	3420
TAACGACAAA ATGGAAGGGC TTAGCTGTAC TTTCAAAAGG AAAGTTTCT AAAAAGGCTT	3480
TGACATCAAC AATTCGCGCC ATATAATAGG GCTCAATGCT TTCTTTAATT TGGCTGTCTT	3540
CTAAAAGAAA GGCTAATGGT TCGTTTTTAT AGATATCGCC CTTGACCCAA TAAACCATCG	3600
AAAAATGCGC AGTAATAAAA TTCCaCAAGC CATTTCGGGC TTCTTGGTTC AAGTAAAACA	3660
TTTCCTTAAT ATGAAATACC TCATCGGCcA CCCAATAGAA AAGTACTCCT AAAGGTTCCT	3720
GATTGGCACC ATAATAGACA GCAGCAGTCC GTTCTTCCTC GTTTTCGAAA CGCCAATATT	3780
CTTCCCAGTT GAAGGCACTA CGAATGAGTG CGCCGTGATT TTGACGAGCG AATCGCGCAT	3840
AGACATCAAA GACATCTGGG TGATCGACAG CTAACCGTTC AATCATTCCT GGAACAGGAA	3900
CGGTTTTTGG TAATTGCGTA TCCCGAATTT TAAAAGATAA CTTGTCGGAC ATAATTTCCC	3960
AACCTTTACG TCGGTAATAA GGGATGTTGT AAGGGAAAAG ATAAGAAATC CACTGTTTAT	4020
CTTGGCGCAT TTCTTCTAAT GCCGTTTGAA TCAAGTCTTT CATTAAACCG TGGTGGCAT	4080
ATTCAAGGATA GGTACCGACA CCAGTAACGC CGCCCATTTT ATATAAAGCG CCATGGATAT	4140
TGACTTCGCA TGGATAGATA GCAATTTGTG AAATTAATTG ATTTTCGTGA AACCAGCCAA	4200
ATACTTTTGA AAGTTCTAAA ATCGGTTGTT TTGATTTGAT AAATGCGCGT TTGTTTTCAA	4260
AACCACTTTC TTCAATATCT GCTTCAGTCA CTTGAAAAAC ATAGGAAAGA AGTTCATTAA	4320
ATTGATCAAT GTGTTcYtCC yCaACAGGCT TTAAnGGTCa GTTGCTTtcC rAAAtCyTG	4380
gTTCcATCCc ATTTTTcCAT CTCCAATAAA TTTActGTTT TTAgtATACC AAATTTTTCG	4440
AAACTCTGTT AGCCTTTTCA AACCAAAAGA ATAAAAGGGA AGTGCTTTAA ATCTTCTTCA	4500
AAAATTGGTA TAATGAAAAG TGCTGATTAT CCTAGACCTT TAAACATGAT AATTTGATAA	4560
AAAAGGCAAC TAAAATATAA CAATTGATTA ATGAAGCTAG CAGGGAGCTA GCTTTTCTTA	4620
AAAGAAATGA GGCAAAAAAT GAACAATAAA GAAATGAAAG CAAGACAAGA GAAAATTCGT	4680
AATTTCTCGA TCATTGCCCA CATTGACCAT GGAAGTCAA CTTTAGCCGA CCGGATTTTG	4740
GAAAAACAA ATACAGTTAG CAGTCGAGAA ATGCAAGATC AATTACTTGA TTCAATGGAT	4800
TTAGAGAGAG AACGCGGCAT TACTATCAAA TTAAACGCAA TTGAATTAAA CTATACAGCC	4860
AAAGATGGTG AAATCTATAC TTTCCATTTG ATTGACACAC CAGGGCACGT CGATTTACAC	4920
TACGAaGTTT CTCGTAGCTT GGCAGCTTG GTAGGGGCTG TTCTAGTTGT TGaTGCGGCG	4980
CAAGGaATTG AAGCGCAAAC GCTAGCaAAT GTCTATTTGG CATTGGATAA TGACTTAGAA	5040
ATTTTACCTG TTATTAATAA AATTGATTTA CCCGCCGCTG ATCCAGAGCG TGTTCCGACA	5100

GAGATTGAAG ACGTAATTGG AATTGATGCA TCGGAAGCTG TTTTAGCAAG TGCAAAAGCA	5160
GGGATTGGGA TTGAAGATAT TTTAGAACAA GTGGTGGAGT ATGTACCAGC TCCATCAGGC	5220
GATATTGAGG CTCCTTTAAA GGCTTTGATT TTTGACTCTA TTTACGATAG TTATCGGGGG	5280
GTCGTTTTAA ACATCCGTGT AATTGACGGT GTCGTTTCGT CTGGGGATAA AATCCAAATG	5340
ATGAGTAACG GTAAAACGTT TGATGTAACA GAAGTCGGCG TTTTTTCACC GAAACCGATT	5400
GCTCGTGATT ATTTAATGGT TGGTGATGTG GGCTATATCA CCGCTAGCAT TAAAACGGTT	5460
CAAGATACAC GGGTCGGGGA TACAGTGA CTGGCCGACA ATCCAGCAGC AGAAGCACTA	5520
CCAGGCTACC GCAAAATGAA TCCAATGGTT TATTGTGGTT TATATCCAAT TGATACGTCG	5580
CGCTACAACG ATTTACGGGA AGCATTAGAA AAATTACAAT TAAATGATGC GGC GTTACAA	5640
TTTGAACCGG AACATCGCA AGCTTTAGGG TTTGGTTTCC GTTGTGGTTT CTTAGGTTTG	5700
CTGCACATGG ATGTTGTTCA GGAACGTTTG GAACGAGAAT TTAATTTAGA GTTAATTACA	5760
ACAGCACCGT CTGTAATCTA TCACGTTAAT AAAACTGACG GAACAACCGT TGTTGTTGAC	5820
AACCCAGCTG AATTTCCAGA ACCAGTAACG ATTGAATCTG TGGAAGAACC TTATGTTAAA	5880
GCGCAAATCA TGGTGCCAAA CGATTATGTA GGAGCAGTAA TGGAATTATC ACAACGTAAA	5940
CGTGGCGAAT TCATTACAAT GGATTACTTA GACGATTATC GTGTAAACGT AGTTTATGAA	6000
ATTCCGTTAT CTGAAATCGT GTTTGACTTT TTCGATAAAT TGAAATCAAG TACAAAAGGC	6060
TATGCATCCT TAGATTACGA AATGGCTGGC TATCGTACCA GCCGCCTAGT GAAAATGGAT	6120
ATTCTATTAA ATGCTGAAAA AGTGGATGCT TTAAGCTTTA TTGTTACCG AGATTTGCA	6180
TTTGAGCGTG GTAAAGCGAT TGTTGAGAAA CTGAAAAAAC TAATTCCACG TCAACAGTTT	6240
GAAGTCCCAG TTCAAGCGGC GATTGGTCAA AAAATTGTGG CTCGTTTACA TATTAAAGCC	6300
TTACGCAAAA ACGTACTGGC TAAATGCTAT GGTGGCGATG TTTCTCGTAA ACGTAAATTG	6360
TTAGAGAAAC AAAAAGAAGG GAAGAAACGG ATGAAACAAA TTGGATCCGT GGAAGTTCCT	6420
CAAGAAGCCT TTATGGCGGT TCTGAAATG GACGACCAAG ATAACGCGAA ATAGCTTGTT	6480
TCCAACGGAA A	6491

## (2) INFORMATION FOR SEQ ID NO: 157:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1014 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

CCTACGAAAA ATATGGCACA AAGAAAGTTG GAAAAATGAC AAACTTGTCTG ATTTTCCAAC	60
TTTCTTCATC TATTCGTGGT TGGCTTGCTG AACTGCATCC ATCACGGCAT AACGGAAGCC	120
ATTTTTTTCT AGAGCTACGA CACCGCGAAT CGTACTTCCT CCTGGAGAAG TCACTTCATC	180
TTTCAATGCT CCTGGGTGTT TTTTCGTGTC TAAGGCTAAG GTCGCCGTTT CTTTTATCAT	240



TTGAGCGACA ATTCATAAG ATAAGTCTCT TGGTAACCCT TCTAAGACTG CTGCATCTCC	300
CAATGCTTCC ATAAAGATAT CAACAAAAGC TGGACTGCAA CCAGCCACAG TACCAAAAGT	360
TTCCAGTTGC GCTTCACGAA TTTCTTTGAC TAACCCTAAA GTAGCAAGAA GTTCCATTGC	420
CAGGTCCTTC GTTTGGCCTT CAATTGTTGG TGCCAAGGCT ACGCCAATCA CGCCTTGATT	480
CACGCTAACT GGTGTATTAG GAATGGCATG AACAACTTTG GCTTCTTTAT TGCCAAGAGC	540
CATTTGTGCT TCTTGATGG TATGACCTGT TGCAACAGAA ATCAATAGAC TTTCGTTGGA	600
AAAATGAGGG GCTAATTCTT GAAGAACAGG TAACACGACC TTTGCTCCTG TCGCAATAAA	660
AATAACCTTA CATTCTTGGA AAAGTGCATA ATCTGTGACA AGTTGAAAGC CCCACTCTTT	720
TTGAAGTGCT TCTGCGGTTT CACTTGATCC GCCTTTCACA TATAAATGTT CCTTCTTGAC	780
TGCCTCAGCT TTCAGTAACC CTTTAATCAT GGCACCGCCC ATACTGCCAG CACCGATAAA	840
GCCAATTTTC ATAGCATTC CTCCTTAGTT AAATAAaGTG TTtGTGkCAA AAGTCAATTA	900
GCTTTtGACA CnAACACTAA CAGTCGkACT TTTTcCgtAC aGtGkTTCG TTAaTTaATT	960
gGTAAtcCAT wACGaCkGCT CaAAcGcGA ATTTTTCaGA AGtCyGaACC AACT	1014

## (2) INFORMATION FOR SEQ ID NO: 158:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16484 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

TTATAAAAGA CCCTTCCTTA AAAAGCAATG TTTCTTCCAA TCAAAAAAAC GAATTATCTA	60
AAAAATGCAA CTTATTTGCT CAAAAAGCA AGCGATGATT GACAAACGGT CGTTTTTAAC	120
TGATAATGAG AATCGTTATC AACTAAAAAT AAAGGAGTTA CACAATGAAA CTTTTAAAAA	180
AGACGGTCCT AATTGGTACA ACCCTTCTTC TTGGTTCATT CTTACTCGCA GCTTGTGGTA	240
ATACGAATAA AGAAGCCAAC AACGCTGACA AAACACATGA AGTAACAGAT ACCTTAGGCA	300
ATAAGTAAC CGTCCCCGCG AAACCCAAAC GGATTATTGC GAGTTATTTA GAAGATTATC	360
TAGTTGCATT AGGAGAAAAA CCAGTGGCAC AATGGACAGT TGGACAAGGC AGCATTCAAG	420
ATTATTTAGC GAAAGAATTG AAAGATGTCC CCACTATTTT CTATGACTTG CCATATGAAG	480
CGGTTCTAAA ATTTGAACCT GACTTATTAT TAATCAGTTC ATCTGCTCTA GTTGAAGGCG	540
GTAAATACAA AGAATACAGT AAAATTGCGC CAACTTATGT AGTCAAAAAAC GGCGAAAATG	600
TCACCTGGCG TGATCAATTG GaAGATATTG CCACTGTTTT AGATAAAAAA GAACAAGCGA	660
AAAAAGTGTT AGAAGATTAT GATACCTTAA CCAAAGGCGT CCAAGAATAT CTTGGCAAAA	720
AAGATGCTGG CAAATCTGCG GCAGTCTTAT GGGTAACCAA CAACCAAGTC TTTATGGTTA	780
GCGATAATCG CTCAAGCGGA ACCGTGCTCT ATCAGGACTT AGGCCTCCAA GTTCCAAAAT	840

TAGTGGAAGA AATTTCTAAA AACGCTACTG CGGATTGGAA, TCAAGTTTCT TTAGAAAAAT	900
TAGCTGAGCT TGACGCAGAC CACATTTTCC TTGTAAACAG CGATGAATCA GCACCTCTTT	960
TCCAAGAAGC AATTTGGAAG AACTTACCTG CTGTGAAAAA TAACCAAGTT CATACCTATG	1020
ATAAAAAAAG TAGTTGGTTA TACAACGGAC CTATTGCGAA TACTCAAATT GTTGAAGATG	1080
TAAAAAAaGC GCTCTTAAAT TAAAAAATG CtTGGATGAA AATAATCATC CAAGCATTTT	1140
TTTTAATTAG CAATTGCCGT TTCTAAACCA ACTTCAATCA TTTCATTAA CGTCGTTTGG	1200
CGTTCTTCTG CCGTTGTTTC TTCGCCTGTC ACCAAGCTAT CACTCACCGT CATAATCGCC	1260
AATGCTTCTA CATCAAATTT TGCTGCTAAA TAATATAACA TTGCCGCTTC CATTTCAATC	1320
GCTAATACCC CGTATCGACC TAAACGAAAG ACATCATCTA AACTATCTTT GTAAAAAACA	1380
TCATCTGACA ATACGTTACC AACATGGGTA GTAAAGCCTT TTGCTTTAGC AATTTCTGTAA	1440
GATTTGAGTA AAAGATCAAA GCTCGCAATT TGTGGAAAT CATATTTAGG AAAATCATTG	1500
CGAATCGCGG AAGAAGGCGT CGCTGCTGCT TGTGCGATAA CTAATTCACG AACGTTGACT	1560
TTTTCTGAAA TAGAACCACA CGTACCAACA CGAATTAAC TCTTCACATC GTAAGAGTTA	1620
ATTAACTCAT GGGCATAAAT ACCTGCTGAA GGCATCCCCA TCCCTGTTCC TTGGACTGAA	1680
ACACGCTGTC CCTTATACGT TCCTGTGTAA CCTAACATCC CACGAACTTG ATTATAGCAG	1740
ACAGGATTCT CTAAAAAAGT TTCCGCAATA TATTTTGCTC GGAGGGGATC CCCTGGTAGT	1800
AAAATTTTAT CGGCAATTTT ACCTGGTTTT GCTTCGATAT GCACACTCAT TCTGGTTTAC	1860
TCCTTTTTTA TAAAGCTGCT AGTGTTGCTT TCACTAATC TTTGAATTGT CCTTTAACGC	1920
GTTCAAGTTGT TTCAACTACT TCTGCATGAT TTAAGCTACT TTGCATACCA GCTGCTAAAT	1980
TCGTAATACA AGAGATGCCT AGTACTTTCA TGCCACTATG CACAGCTACA ATGACCTCCG	2040
AAACAGTTGA CATGCCAACT GCATCCGCAC CCATCGTCCG TGACATTCGA ATTTCTGCAG	2100
GCGTTTCATA TGTCGGGCCA GAAAAGCCCA TATACACGCC TTCTTTTAAA TCAATATTTT	2160
GTTCGCAGC CACTTTTTTT GCCACTTCCC GATATTCTTG CGTGTAAAGCA TGACTCATAT	2220
CTGGAAAACG CGGCCCGATT TCTTCATCAT TTTCACCAAT TAACGGATTA TCGCCTGTAA	2280
AATTAATGTG GTCATTGATT AACATCAAAT TTCCCGGTGT ATATGTTTCA TTTACGCCAC	2340
CAGCTGCGTT GGTACAATC ATTGAATGAA TGCCATAAGC AGCCATCACA CGAACGGGAT	2400
AAGTGACCGT TTGCATCGAA TGTCCTTCAT AATAATGAAA ACGCCCTTGC ATCGCCAAAA	2460
CTTTTTTTCC AGAAAGCGTT CCGTAAACTA ATTGTCCAGC ATGTCCAACG ACTGTTGAAA	2520
CAGAAAAATG GGAATTTTCA GAGAAAGGAA TAGCAATCGC ATCGGTTATT TCATTAGCTA	2580
ATTACCTTAA CCCAGAACCT AAGATTAGAC CGAAGTCTGC TTGTTGCACC CCTTTTTCTT	2640
TCAAGaCTC AGTTGTTTCG TTAAATGAT TCGTTAACGT TGTCATGTGT TGCCTCCTCG	2700
AATTATTGTA ATTGATTTAA AAAGCTTTGA CCGTTTTCCG TTGCAGGAAC TTCAAAGTTT	2760
TCAGCAATCG TTGCTGAAAT ATCTGAATAA AATCCTTGTG GCAAGCTGCC TTGTTTTGTC	2820

ATTTTTTTAC TATAAGCCAA TAAAGGCACA TATTCACGTG TGTGGTCTGT TCCTGGGAAA	2880
GTTGGATCAT TCCCGTGATC TGCTGTAATT AAAAGCAAAT CATCTTCGGC CATAGCTGCC	2940
ATAATCTCTG GCAAACGACC ATCAAAATCT TCGATGGCTT TTGCATAGCC CTTTACATCA	3000
CGACGGTGAC CATAACGCGC ATCGAAATCC ACTAAGTTAG TAAAACTTAA ACCTGTGAAC	3060
TCTTTGTTCA TGACCGTTAA TAATTGATCC ACCCCATCCA TATTGCTCTT CGTGCGGATA	3120
GCTTCAGTAA TTCCTTGTC ATTAAGATG TCGTTAATTT TCCCAACAGC GATCACATCA	3180
TTGCCGTTTT CTTTCAACGA ATCTAAAACA GTATGGCCAA AAGGATCTAG AGCATAGTCA	3240
TGACGGTTAC TGGTTCGAGT GAAATTACCT GGTTCACCAA CACGCGGCG AGCGATAATC	3300
CGACCAATCA TATAAGGTTT GTCTTTCGTA ATGTCACGAA CATATTGACA AATTTTATAC	3360
AATTCTTCTA ATGGAATAAT GTCTTCATGC GCCGCAATTT GTAAAACAGG GTCTGCTGAA	3420
GTATAACGA TTAGATCCCC AGTTTTTATT TGATGTTTCG CATAATCATC AATGACTGCT	3480
GTCCCGCTAT ATGGTTTGTT ACAAACAAC TTTGCCCCAG AAAAATCTTC AATTTGTTTC	3540
AGTAATTCTT CAGGAAAGCC GTTTGGGAAT ACCCGAAAAG GTTTTTTGAT GTTCAAGCCC	3600
ATAATTTCCC AGTGACCGGT CATAGTATCT TTTCCAACAG AAATTTCTTC TAATTGGTA	3660
GCATAGCCAT CGTGGTCTGC CACTGCTTTG ACACCTGTTA AAGGAGCAAT AGTTCCCAAT	3720
CCAAGTTTTT CTAATGAGG AATGGTATG CCCGCTTCTT TGGCAATATG ACCCaGTGTA	3780
TCACTGCCGA CATCTCCAAA TTTTCAGCA TCTGGTGCTT CACCAATACC GACTGAATCt	3840
AGTACGATTA AATGAACACG TTTAAACACA CTTGCTTCCT CCTCTAAGGT CGTAAATACG	3900
TCAAGAATGA CAATTATTTT ACGCTAGCCT TTTAAATAGT CTCGAGAATC AAACAGAATA	3960
ATCTATTTAT TCTCAGCTCT TCAATTATAC CATGAATCCG CTATTATTGC CTGAAAGAGC	4020
ATTAATAATT TCACTGATTC CTTTGATTTT AGCAGGGCTT AAAAATTGAA AGAAAAAAG	4080
TCTCCACCG TAATTAGGAA ACTTTTTTAT TGTTTTTTTC GAATAGCCTG CTCCTTAAAA	4140
AGAACACGAC TCTTTATCCA ATCGAAACTT TTAACACTCC AAGCGGTCTC CCTCTTTTGG	4200
CACATTTAAA TACCAGATAA CCGCTTCTGA TAAATCACA CCAATGCCAA TGGGCGGATA	4260
AAAATAAATA CCGATATATC CGATTGCGAA AAGACAAAGA TGAATGATTT TTGTTGCCTT	4320
ACTATAAGAT TTCGATTTTC TTAACCTAGC TGCACCTTGT GGATCTTCA CTACATTGTC	4380
TTCATAAATG ATCCTTGTTA GCAAGAAGTA GGCAAAATCC CaAACCGTAT ACACAACAAG	4440
ATAAAAATAA ACAGGTGCCC GCATCCACCA AAATTCACCT ACCCAAGCAG TTGCTACCGG	4500
AATGAAGGCC ATGGAGAATA ACCAAAAATT ATTTGCCCAA AACGCTTTTT TAGAAAACCA	4560
ATGAGTTACG GACATTATGT AATGATGATT GTACCATGCG ATACAGATAA AAATAAACT	4620
GACAATAAAC GATAAAAAAT ATGGGATATT TTCTATCAAA CCTGCCAAAA CACCCACTTG	4680
AGGAGTTTTA ATTTCTAAAA TCATGATGGT CATGATGATG GCAATGACTG CATCACTAAA	4740
TGCTTCTACG CGTGTTTTAT TCATTGAACG TCCTCCACTT TTCCTATGCT TACTACAAAC	4800

AAAAGTATAG CAGAAAAACC AACGATTTTT GAAAAAAGC TGCTAGACCT GTTTGTTTTT	4860
TTTCAAATT TCAACACTTC CTTCGAATTG TAAAGAACAG CTAAAAGAAA AmAGAAAAGT	4920
tACCTTTATC AACTCGCTTT TTTACAGTGA GGCAGTGCCT AATGCCGCTT GTGTTCAACT	4980
GATTATCCTA AGAAAAGAAA TGGCCCCCG TCGCTTTGAC GGAGGGCCTA TTTCTTTTct	5040
TATTTCTGTTT TGACGTAATT TTTCCCGTTC GCTTTTGGTC CCGATGATTT TCCGAAGAAC	5100
ACAACTAGGA CTAAAATGGT TAATACGTAT GGTGCTGCTT GTAGATAAAC TTTTGGAATT	5160
TCTGAAATAA TTGGTAGGTT TTCCCCAGCA ATACTTAAGT TTTGTGCAA GCCAAAGAAT	5220
AAGGCTGCGC CCATTGCCCC AAGTGGATTC CATTTACCAA AGATCATTGC GGCCATTGAA	5280
ATAAACCTT GTCCTGCAAT TGTGGTAATC CAAAACGAC CCGCAATACT TTGAGCAAAG	5340
ACAGcTCCGC CCATACCGCC TAAGAAGCCG GATAGTAAGA CACCAGAATA ACGCATTAAG	5400
TAGACATTAA TTCCCAATGT ATCTGCCGCT TGTGGATGTT CCCCCACAGA ACGTAACCGT	5460
AAGCCAAAAC GTGTTTTATA AATGACAAAC CAAGCTAAAA TCGCAATCAA AATTGCTACG	5520
AAAGCTGGCA AAGAGGTTTT TGCAAAGAAT AGTTCACCAA TCACTGGAAT TTTTGACAAT	5580
CCTGGGAAAG AAAAATAACC AAATGATTGT GTAATCATAT CTGTTTGTCC TTTTGATAA	5640
ATAACTTTTA CTAGGAAAAC CCCTAGCGCC GGTGCCATCA AGTTAATGAC AGTACCACTG	5700
ATAATATGAT CTGCTCGTAA GTTAATAGTT GCAACGGCAT GTAAGAGAGA AAAGACAATC	5760
CCGACAACGC CTCCTACTAA ACAAGCAATC CATGGTGTGCG CCGCGCCAAA TGTATCAGCA	5820
AACGTCAAGT TAAAAACGAC CGAACTGAAA GCACCCATTA CCATAATACC TTCCAAACCA	5880
ACGTTAACCA CACCACTTCG TTCTGAGAAC GTTCCGCCTA ATGCTGTCAA GATGAGCGGA	5940
GTTGCATACA CTAATGTTTG CGTGATGATT GAGGAAATTG TTGCTAAATC CATTAGATGT	6000
CTCCCCCTC TTGATTGTTT TTGTCTGGCG CCGCTTCAAT TGTTTCTACA ATCGCTACTT	6060
CTTTTTTACT ACCAGACGCT TTTGCTAACA AGAAACGAAT TAAATAACTA ATTCCCACAA	6120
AGAAAATAAT CGCTGCAATA ACAACATCAA CTAATTCGGT TGGTACACCA GCTATCAACG	6180
GCATCCCTTG TCCACCAAGT TTTAACATCC CGAACAGTAA TGCGGAAAGT AAAATCCCGA	6240
TTGAGCTACC TGCTCCTAAT AGAGACACGG CCATACCATC GAACCCAATG CTCAACGATG	6300
TGTTTTGAAT AAAGAAGTTT TGAAAGTTC CCAATCCTTG GACCGCACCA CCTAAGCCAG	6360
CTAGCGTTCC TGAAATCATC ATGGATAAAA TAATCGTTCG CTTGCTGCTC ATTCCAGCAT	6420
ATTGAGAAGC AAATGGGTTT ACCCCACCG ACCGAATTTC AAAACCAAGC GTTGTTTTTT	6480
TCATCAAGAA CCACACAAGA ATAAGGAAAA TGAAGGCCAA GAAAATGCCG ATGTTCAAAC	6540
GAGAGCCATC ACTGATACTT GTTAAAAACG GTGAGCGTAA TGATGCATTC GCACCAATCA	6600
TTTTGGTGAT CCCTTTGTTC GCCATAATAC TTGGTTTCAT GACATTATTA GCAATATGAT	6660
TTCTGTGTA TAATAAAATA TAATTCAACA TAATCGTAAC AATAACTTCA CTCGTCCCAA	6720
AATAGGCACG CAAGAAACCT GGAATGGCTG CTGCTAAAGC TCCTGCAACT GCCCAACAA	6780

TAATCGCCAC	CGGCAACGCA	ATCATCCGGG	GTGCATCAGG	CATGGATAAA	GCGACCCAAA	6840
TACTAGCCAC	CCAACCACAT	AAGGCTTGTC	CAGAAAGCCC	GATATTAAAG	AATCCAGCCG	6900
AATTAGCCAC	CGAGAATCCT	AAAGCGGTAA	AAATCAACGG	TGCTGCGGTG	ACAAAAATCT	6960
CACCAATACT	TTTCTTTCCT	TGGAAAGCTG	TTTCAAACAT	CGCTTGATAC	CCAGCCACTG	7020
GGTCATAGCC	AAAGATAACC	ATAATAATTG	CCCCTAAGAG	GAAACCCATA	ATAACAGATA	7080
ACACAGGGAC	TAAAAATATTT	CTTATTTTTT	CAGAACGATT	ATTCAACAGT	TTCACCTGCC	7140
TCTACTTTTG	ATAACTCAGC	ACGGGCTTCT	TCCAATGAGT	AGCCAGCCAT	TAACAAGCCC	7200
AATTCATTTT	CTGTCGTTTC	TTTCGGATCA	ACAATTCCGA	CAATTTCCCC	CGCATGAATA	7260
ACAGCGATAC	GATCAGAAAC	GTTTAAATTT	TCATCTAATT	CAAACTTAC	TAATAAAACA	7320
GCTTTATTTT	TATCTCTTTG	TTCAATTAGT	CGACGATGAA	TGTATTCAAT	AGCTCCGACA	7380
TCTAAACCGC	GTGTTGGTTG	AGAAACAATC	AACAAATCTG	GATCGCGGTC	CACTTCTCGG	7440
GCAATAATCG	CTTTTGTCTG	ATTCCCACCT	GAAAGCGCCC	GAGCTGGTAC	TAATTCATTA	7500
GGAGTCCGTA	CATCATATTC	TTCAATTAAT	TCACGGGCAT	GTTTGTTAAT	CATTCGATAG	7560
TTTAAAATAC	CTGTTGAACT	GTATGGTTTT	TGGTAATACG	TTTGTAGGCC	AATATTTTCA	7620
GCCAATGTCA	TCTCTAATAC	CAAGCCATAT	TTATGTCGGT	CTTCAGGAAC	GTGGGCGACA	7680
CCCATTTCCG	TAATTTCCCG	AGGTTTGC GA	TTAGTAATCT	CTTTTTCATT	CAAATGAATG	7740
GACCCGTTTT	CGACTTTGGT	TAATCCTGCT	AAGGCTTGGA	TTAATTCTGT	TTGTCCATTG	7800
CCATCAATCC	CCGCAATTCC	AACAACTTCC	CCTGCTCGGA	CATCTAAGCT	AAGATTTTTC	7860
ACTGCATCTA	AGCCGCGGCT	TTCTTTCACT	ACTAAATCTT	TAATTGATAA	AATTGTTTCT	7920
TTGCGTTGCG	CTGGTTTTTT	CTCTGTTTTG	AAGGAAACAG	AACGACCCAC	CATCATATCG	7980
GCTAACTGTT	GCGAAGAAAC	GTCTTTCACA	TTGACGGTGC	CAATACTTTG	CCCCCGGCGA	8040
ATAACCGTAC	AACGGTCCGC	CACTTTTTTA	ATTTCACTA	ACTTATGTGT	AATTAAAATG	8100
ATCGATTTGC	CTTCATCGGT	TAGTTCTTTC	AAAATAATCA	TTAATTCATC	AATTTCTTGT	8160
GGCGTTAATA	CTGCGGTTGG	TTCGTCAAAA	ATTAAACGT	TCGCACCACG	ATATAAGGTC	8220
TTTAAAATTT	CAACCCGTTG	TTGCATACCC	ACAGAAATAT	CGCGAATATA	GGCATTGGGA	8280
TTTACATCCA	ACCCGTATTG	CTTAGACACT	TCTTCAATCT	CTTTTTTCGC	TTTTTTCCGG	8340
TCTAAAATAC	CGCCATGGGT	AGGTTCACTA	CCTAAGATAA	TATTTTCAGT	CACAGTAAAG	8400
GCTTCAACTA	ACATAAAATG	CTGATGGACC	ATTCCGATTC	CTAAACGATT	GGCTGCAGTA	8460
GGACTTGTA	TGTTGACTTT	TTGCCCATTC	ATATAAATCT	CACCTGAGGT	CGGTTCCAAT	8520
AGCCCTGATA	AGATATTCAT	CAAGGTCGAC	TTGCCTGcAC	CATTTTCACC	TAGCAGTGCG	8580
TGGaTTTCAC	CAGGTTTTAC	TTGTAAATCA	ATTTGGTCGT	TGGCTTTGAA	CGTCCCAAAT	8640
TCTTTGGTGA	TTTTACGCAT	CTCAATTACG	TTTTCTTCTT	GAGACAAGGT	CTATTCAC TT	8700
CCTAACAAAA	TTATATTGGC	AATACTTTAG	AGGGTAATTT	TGACCCTTAA	CTAAGTAGTT	8760

GAAGAACTTC	AACTATTTAG	TTAAAAATCC	AAACATCATA	AAGGGTCCC	TTGGCAGCCG	8820
TCGCTGCCAA	GgACCACTTT	ATTTTAGAA	GGAACAATAA	TGAGCCCTTC	TTTTTTCATT	8880
TAGGAGCTTA	AATTATTCTG	GTTTTCTGG	GACTTTTACG	TCACCTGAGA	TTACTTTATC	8940
TTTTGCTGTT	TTAACAGCTT	CTTTTGTTTT	GTCGTTTAAA	TAGCCGTCTG	TTAAGTCAAC	9000
GCCACCATCT	TTAATCCAT	AACTAAATG	TTGCCACCA	GGGAATTGT	CTTCTAACGC	9060
ACGGTTGGCA	ATATCTTGAA	CCGCTGTGCC	GACACCTTTA	AGCGTTGAAG	TTAACGTGAA	9120
GTTGTCTTCT	TTGCCGTCTT	TTGTTTTGTA	CTTGCCATCA	GCATCTTGAT	CGCGGTCAAC	9180
GCCGATTACC	CAAACCTTGT	CGCCAGAACC	TGATTCATTC	AAGTCTTTTG	CTTCTTGAA	9240
GACCCCTTGT	CCAGTCGCAC	CAGAAGCATG	GAAGATGATA	TCAACGCCGT	TTTGGTACAT	9300
TGCAGCAGCT	AAAGCTTTCC	CTTTGGCAGG	ATCAGCAAAT	GAAGCCGCAT	ATTCGTATC	9360
AACAGTAATT	TCTTTACCTA	ATTCTTTCGC	AGCATCAGCC	ACACCTTTTT	CAAACCCAGC	9420
TTGGAACGG	TCAATTACGA	CCCCTTCTTC	ACCACCAACA	AAACCGACTT	TGTTTCGTTTT	9480
TGTTTCATTT	GCAGCAGCAA	CACCGGCTAA	GTAAGCTGCT	TCATTATCTC	TAAATGTTGC	9540
AGAAACGACA	TTCTTTTTGC	CATCGATTG	ATCATCGATT	AAAACAAAGT	TTGTATCAGG	9600
GTTGGCATCT	GCTGCAGAAG	AAATTGCATC	TTTTAGCAAG	TAGCCAATAC	CAAAGATTGT	9660
GTTGAATTTA	CTTGATACCG	CTTGGTCAAT	ATTGGTTGTA	TAGTCAGCTG	CATCATTCGA	9720
TTGAATATAT	GCATACCCTT	TTGAACCTTC	TGGTAAATCA	TGTTCTTTAC	CCCAAGCTTG	9780
CAATCCTTCC	CAAGAAGATT	GGTTGAACGA	CTTGTATCC	ACGCCGCCTG	TATCTGTAAT	9840
GATTACAGCA	CTATGCGCTG	CATCCCCTTT	GCCACCGCCG	CTTTCAGCGG	TTTTGCCTTT	9900
GCCACCACCA	CATGCTGCAA	GTGAACTGA	TAAACCTAAT	GCAATCAAAC	TAAAACCGAA	9960
TAATTTTGCT	TTTTTCATTT	CTGTAAAGCC	TCCTAAAAAT	AGTTTAATGT	ATTTAAACAA	10020
CTAAGTTGTT	TCTAAGGTTG	GTCTGGGACT	TTTACGTCAC	CAGAGATGAT	TTGTTCTTTG	10080
GCTTCTTTAA	CTGCCGTTTT	GGCTTGATCG	CTTAGGGCTT	CTGTTGTAA	ATCAACACCG	10140
CCATCTTTCA	ATCCATAAGT	TAAATGTTTCG	CCGCCAGGGA	ATTTATCTTC	TAATGCAAGA	10200
TCAGAGATTT	TTTTCACAGC	AATATTTACA	CCTTTAATGG	TTGAAGCTAA	CGTTAAGTTA	10260
TCATCCTTGC	CATCTTTGGT	TGTATATTTA	CCGTCTTCGT	TTTGGTCACG	GTCAACACCA	10320
ATCACCCAAA	CTTTATCTTT	TGAACCTGTT	TCATTTAGCG	CTTTGGCTTC	TTGGAAGATT	10380
CCTTGTCCTG	TAGTCGCCGC	TGCGTGATAA	ATAATGTCGG	CACCTGCTTG	ATACATAGAA	10440
GAAGCTAAGG	CCCGTCCTTT	TGAAGCGTCT	GCAATGTCC	CTGCATAAGT	AGACGTAATT	10500
TGAATATCTT	TGCCTAATTT	TTTGCCAGCG	TCTGCTACAC	CTTTTTCAAA	ACCAGCTTGG	10560
AAACGGCCAA	TAACAGGTCC	TTCAACCCCG	CCGATAAATC	CTACTTTGTT	GGTTTTTCGTA	10620
GTGTTAGCTG	CTGCCACACC	TGCTAAATAC	GCAGATTCAT	TATCTCTAAA	TGTTGCAGAA	10680
GCGACGTTGT	TTTTCCCAT	AACTGTATCG	TCGATTAGAA	CGAAATTCGT	TTCTGGGTTG	10740

GCGTCCGCTG	CATCTACTAC	TGCATTTTTT	AGTAAATAGC	CAATGCCAAA	AATCGTTTTA	10800
AATTGACTTG	AAATGGCTTG	GTCAATGTTT	GATGTATAAT	CAGATGCTTC	ATTTGATTGA	10860
ATATACGCGT	ACCCTTGAGG	TCCTTCTGGA	AGTTTGTGCT	CTTGCCCCA	CTCTTGCAAT	10920
CCTTCCCACG	CAGATTGGTT	AAAAGATTTG	TCATCAATGC	CGGCTGTATC	TGTGACCATT	10980
ACGACACTAT	GTTGTGGATC	ACCTTTTGTG	TTGGCGTTCT	TTTTGCTATT	GCCGCCACAT	11040
GCAGCTAACC	CAATGGTTAA	GGCAACTGCT	GCCAATCCTA	GACCTAAAAA	TTTTGCTTGT	11100
TTCACTTTCC	GAAAAGCCCC	CTACTATTTT	ATCTATTATT	TCATTGCGAT	TTATGCTATT	11160
GAAATCTACT	TTGAGCCCAA	CGAAACCTTA	TGTCCTTTTT	TTAAAAATCT	TTATCTGTAA	11220
ATGAATAGGG	CAATAATGCC	TCAACTGTCA	TTGTTTTAGT	CGCGCCATCT	TTACCAACTA	11280
GCGTGACCGG	CATTTGTGGA	TCAAAGAATT	CAGCCATTAC	TTGACGACAA	GCGCCACATG	11340
GAGAAATGGG	TGCCTCTGTA	TTACCTGCTA	CAACTAAATG	TTGGAAGGTC	CGTTCATCTT	11400
CTGAAATGGC	TTTAAAAATC	GCCGTTCTTT	CTGCACAATT	TGTTAAACCA	AAGGAGGCAT	11460
TTTCAATGTT	GACTCCTTGA	TACATTTTCC	CGTCTGCAGT	CACTAAGCAA	GCGCCCACAG	11520
GAAAATTAGA	ATACGGTACA	TAGGCGTG TG	CCAATGCTTC	TACGGCACAA	TCAAGCCATT	11580
CTTGTTTTAC	TGTCATCTTT	TTCGCCACAC	CTTCGTTTAA	TTTGAAATAC	GAATTAATAC	11640
CCTTCGCCAG	TTGCGCCAGA	AACAATTGCT	ACACCCGTAC	TTGCACCGAT	GCGTGTTGCC	11700
CCCGCTTCAA	TCATTGcTAA	CGCTTCTTCT	GCATTATGCA	CGCCACCAGA	AGCCTTAACG	11760
CCCATGTCAG	GACCAACTGT	CTCACGCATT	AAACGAATGT	CTGCTACTTT	AGCGCCGCCT	11820
GTTGAGAAGC	CAGTAGATGT	TTTAACGAAA	TCGGCACCTG	CTTCTTTGGC	TAATTCACAA	11880
GCTTTAACTT	TTTCTTCGTC	TGTTAATAAT	GCTGTTTCAA	TAATCACTTT	GACTAACGCT	11940
TTTCCTTTGG	CTGCATCTAC	TACGCCTTGA	ATGCTTTGAC	GTACGTAATC	GTATTGtTGT	12000
GATTTTAATG	CACCAATATT	AATCACCATA	TCCACTTCAT	TCGCACCATT	TTTAATCGCA	12060
TCAGCTGCTT	CGTATGCTTT	AACTTCTGGT	GTATTTGCCC	CCAAAGGAAA	GCCAATAACG	12120
GTACACACAG	CGACGTCGGT	ATCTGCTAAT	TGCTCGCTAG	CAAAAGCGAC	CCAACAAGGA	12180
TTGATACATA	CAGAGAAAAA	GTTGTACTCC	TTAGCTTCGT	CAATAATTTT	TTGGACAGCG	12240
GCTTCTGTCTG	CTTCTGGTTT	CAAAATAGTG	TGGTCAATCA	TGCGGTTTAA	TTCCATGATG	12300
AGAATTCCTC	TTTTCTTTTT	GTTTTATTCC	GTAATAATAT	CGTGAATCAA	AATAGGCTCT	12360
TCGCCGGTTG	GTCCAATTTT	AATGTTTTTA	TAAAGCAATT	GTTCGACTTC	TGTCACGTCT	12420
TCACGATTGG	CATAAATGGT	TAGTAATGAT	TCGCCTTTTT	CAACTGGGCT	CCCTACTTTT	12480
TTGTGAAGTT	TCAAGCCAAC	TGCAAAATCA	ATCGTATCTT	CTTTCGTAGC	ACGACCAGCG	12540
CCCAGCAACA	TTGCGGCAAC	GCCAATTTCA	TTGGCAACGA	TTTTTTGTAC	CACGCCTGTT	12600
TCTTTGGCAG	GTAAATCAAA	TTGATAAGTT	GCGGTAAATA	ATTTTCTCTG	CTGGTCAACA	12660
ACGGAACTGT	CGCCGCCTTG	ATTTTCAATC	ATTTCTTTAA	ATTTTGCGAG	CGCTCGACCG	12720

CTTGTTAATG CTTCTTCTAA TAAAGCCCGT GCTTCGTCTA ACGTCTCGGC TTTTITAGCT	12780
AAAACGACCA TCTGACTTCC TAAGACATAA CACATCTCCA CTAAATCTTC TGGCCCATTT	12840
CCTTGCAAGG TCTCGATTGC TTCAACGATT TCTAAACTAT TTCCGATGGC CTCCCCTAAA	12900
GGTTGTGACA TGTCTGAAAT TACTGCCATC GTTTGGCGAT TTGCCAGTTT TCCAATGCCA	12960
ACCATTGTTT CCGCTAATCG ACGAGCACTT GCTAGGTCTT TCATAAAGGC CCCTTCACCA	13020
GTAGTCACAT CTAAGACAAT CGCATCTGCC CnTGCGGCAA TTTTTTACT CATAATGGAA	13080
CTTGCAATTA ATGGAATTGA GTCAACAGTG GCGGTCACAT CGCGTAACGC ATAAAGTTTT	13140
TTGTCTGCTG GTGCTAGATC ACCTGACTGA CCAATTACTG CCACTTTTGT ATCGTTCACC	13200
AATTGAATAA AACGTTCTTC TGTC AATTCA ACTTTAAAAC CAGGAATCGC TTCTAATTTA	13260
TCCAAcGTAC CGCCCGTGTA GCCTAAGCCA CGACCAGACA TTTTGTCTAC ATTGACCCCG	13320
ACACTAGCAA CTAACGGCGC TAAATAAGC GTCGTAGTGT CGCCTACGCC ACCTGTAGAA	13380
TGCTTGTCGA CTTTGATTCC TTCTATAGAA GAAAGATCCA CAATTCGCC TGAATGGACC	13440
ATTGCTAACG TCATTGTGT AATTTCTTCG TCTGTCATAT CTTGGTAAAA AACAGCCATT	13500
AAAAAGGcAC TTACCTGaTA ATCTGGaATT gAACmTCTG tGtATCCGTT GATAATAAAT	13560
TGAATTTCTT CGCTTGTTAG TTGATGCCCG TCaCGTTTCT TTTCAATAAT GTCTACCATT	13620
CTCATGAATT CTTTCCCTCC GATTTCCCGa TAATGTATTA TACAACAAAA TCGTTCAGCT	13680
TTAAACCATT TAGTAAACCA GTTTACTATT ATTTTTTTCT TGTGAAAACG CAAACAACAC	13740
GAACATTCTA TTTCTAAGCT GTTTTTTAAA TCGGTTTTGT GCTTTTTCTA GCAATAAACG	13800
TCGCATCGAA AACTTTATTT GGAACTTTCC CTGTTGGATC TGCAATAGCT TCCACCAAAA	13860
ATTTTGCTGC ATAATAGCCA ATATCAAAAC TCGGCTGATA GACAGTTGTT AAACTCGGCA	13920
CCAAATATTC GGAAATTTCC AAGCCATCAA AACCACCAC CGAAATATCG TCAGGAATTT	13980
TCTTGCCCGC TTCTTCAATG GCTTGATAGG CCCCCATCGC CATTTTCATCA TTGCCACAAA	14040
AAATTGCTGT GATTTCTTCA TTTTTCAAAA CTTCTTTTCG CGCAGCATAG CCGCCATGCA	14100
AGTTCAATTC CCCAGCACAA ACATAATGCG GACGAAAAGG TAAATTATTG TCCTGTAACG	14160
CATGCTGATA CGCCGTCACT CGTTCCGTCA AACGATAATA GCCGGCACTT TCACGCAACA	14220
TGCCAATATG CTGATGCCCT TGTTGAATCA AATGATGGAC CGCTTGATAA GCACCTTCGT	14280
ACTCTTTAAC GATTAATCGT CCACTATCTC GTGGATTAAT CCCACGATCG ACAAGAATGA	14340
CAGGAATCGG ATGTTCTCGT CGGTCTTTTA AAATATGATT TTCTGGCAAA ACATTCGGCG	14400
TGGCAAAAAT AATGCCATCC ACCGACCGAT GAATCAACTC GCTAACATAT TGCATCGCTT	14460
GCTCGGAATC ATGCTTTGAA TTGCATAACA AAATCATGTA GCCTAACGAA TTTAAATATG	14520
TTTCTACACC TTCAATCACT TTTGAGAAAA AGAAATCCGT CACATCTGGC ACAATCATAC	14580
CAATGGTTTT GGAATGACGG TTGATTACAT TCGAAGCAAA GTAATCTGGT TTGTAATGAT	14640
TTTCAGCGAC TACCGCCAAC ACCTTGTTGC GGGTTTTTTC ACTGAAACGA CTCCCTTTAT	14700



897

TATTCAGAAT TTGTGAGACG GTTGTGACAG AAACACCAGC CATATCAGCA ATTTCTTTAA	14760
TCGTTAGTTT CATCGCGCCC CTCCCAACGT CTCCTTTTGC ATTGACAGAC ATTTGTAGAC	14820
GTGTTACAGT CTTCAAAATA CTTATCGAAT CTTTACATT TTAACAAAAA AAAGACAAAA	14880
GAAATAGACT AAAAAGATTA AAAAACAAAA GAAATGGAG GCTATTTTTT TCATGGAAGA	14940
ATCCCGTGTC CGTCAATTAC CAAAAATCGA GCTGCACTGT CACTTAGATG GGTCGATTCTG	15000
ACCAACCACT TTACGCACCA TTGCTGAAAA ACAAACATT CCATTACCCC AAGACGAACA	15060
AGCCTTGAAA GAGTTAGTGG TTGCCCCTGA AAAATGCACC GATTGAATG ATTATTTAAC	15120
TCGGTTCGAC TTCGTGCTCA CCTGTCTACA AACAGCTGAG GCTTGCAAG CAGCGGCGTA	15180
TGACGTGATT AGCCAAGCTG CCGAAGATGG CGTGGCTTAC ATAGAGGTGC GTTTTGCGCC	15240
TTCCCAACAT ACAGAAAAAG GCTTACGGTT ACCTGAAATC GTGACTGCTG TTTTAACTGG	15300
CTTAAACAA GGCGAAGAAG ATTTTGGTGT AAAAAGCAAT GCACTGTTAT GCGGCATGCG	15360
GCATGACCAA CAACAAGCCA TAGAAAAAAT CGTTCACCTA GCGCATGATT TCAGAGAGAC	15420
AGGCGTTGTC GGCTTTGATT TAGCTGGCAA TGAAGTCGAT TTCCCTCCAT ACACATTTGA	15480
AGACGTATTA GCTTTAGCCA ACCAATTGTC AATTCCATTA ACACTCCACG CTGGGGAATG	15540
CGGCTGTGGC AAAAATGTCG CTGATGCTGT CACTTTAGGT GCTACTCGGA TTGGTCACGG	15600
TATTGCTTTG AAAGATACAC CCGAATATTT GGCGCTCTTA AAAGAAAAAA AGGTTCTTTT	15660
AGAAATGTGT CCAACTAGCA ACTTCCAAAC AGGAACCGTC AAAACATTGG CTGAATACCC	15720
TTTTCAACAA TTTATAGAGG CTGGCTTAGC CGTCTGCATC AATACCGATA ATCGAACTGT	15780
TTCAGATACT ACGTTAACGA AAGAATTCAT GAAATTGGCC ACTTGGTATC AATTAAGTTA	15840
TGACGAAATG AAGCAACTCA CTAAAAATGC GCTTGCAGGT GCTTTTTTAT CGCCAGACGA	15900
AAAAAACTG TTGAACCAA AAATTGATCA AGCGTATCTA TTCTAAAAAA GAACGCCACT	15960
AAAGTGGCGG TTCTTTTTTA CTCTGCTTCT TTAACATTCG TAAAATATAA TAGCCTTTGT	16020
CTTTTGTAC AATTTCAACA TTGCCAAATA CTTAGCCAT TTTCTTTTCT GCACTAGGTG	16080
CACCTTGTTT CTTTGAATA ACCACCGTTA GTGTGCCACC AACTTTCAAC AACGGAAGAG	16140
CCCCTGTAA AATGCCATGG ACAACTTTTT TGCCTGCACG AATGGGCGGG TTGCTAACAA	16200
TCGCCGCGTA GGTCGTTTCG TTTAAGGTTT CATAACATT GGAAGAATGG ATATCCACAG	16260
TAGTAATTTG GTTACGTTGG GCGTTCATTT GAGCTAGCCC AACTGCCCCG TGGTTGACAT	16320
CCACCATTTT CACCAGTCGA CCAGTCGCCG CTGCTAAGGA CAACCCAATT GGTCCATAGC	16380
CACAACCAAC ATCTAACAAA CGGCCTGCTG GTAAATTTTC CCATTCAAAG GCATCAATCA	16440
ACACACGAGA ACCATAATCC ACTGTTCTCT TTGAGAAGAC CCnG	16484

(2) INFORMATION FOR SEQ ID NO: 159:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8160 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

CCGAGTAGTA CCAGAGAGGm AnTCCATGTT wAAAAcATTA CTTCTTTmCA ATCTAGTTAC	60
AGTAATATG <sub>a</sub> AATTTTGTG AAGTGTTTTT AATTCGCCAG AAAAAAGGT ATACTTGTTA	120
TAATrACAAA ACTAAAAAT TACATTTACA TATAAGTTAA AAAAGGAAAG TTGGGGACGT	180
ATCAATGAAG AAAGAATCAA TGTCACGTAT CGAAAGAAG AAAGCACAAC AAAGAAAGAA	240
AACGCCAGTA CAATGGAAGA AGAGCACTAC TTTATTCAGC TCGGCGTTAA TTGTTTCATC	300
TGTAGGAACG CCCGTTGCGT TACTACCAGT GACTGCTGAG GCAACAGAAG AGCAGCCAAC	360
AAATGCGGAA GTTGCCCAAG CACCTACTAC GGAAACTGGC TTAGTAGAGA CACCAACAAC	420
AGAAAcTACG CCAGGAAC <sub>T</sub> A CGGAACAACC GACAACGGAT TCGTCAACAA CGACTGAATC	480
GACAAC <sub>T</sub> GAA TCATCAAAAG AAACACCAAC AACACCAAGT ACCGAGCAAC CAACAGCTGA	540
TTCAACTACA CCTGTGGaAT CAGGaACGAC TGATTCTTCA GTAGCAGAAA TTACGCCAgT	600
AGCTCCTTCA GCAACCGAGT CTGAAGCAGC GCCTGCGGTT ACACCCGATG ATGAAGTAA	660
AGTACCAGaA GCTAGAGTAG CTTCTGCGCA AACTTTTTCA GCGTTATCAC CGACGCAAAG	720
TCCTTCAGAA TTTATTGCCG AGTTAGCTCG TTGTGCACAA CCTATTGCGC AagCCAATGA	780
TTTATATGCA TCAGTGATGA TGGCTCAAGC AATCGTTGAA AGTGTTGGG GAGCAAGTAC	840
GCTATCTAAG GCACCAA <sub>A</sub> CT ATA <sub>A</sub> CTTATT TGGGATTAAA GGCAGCTACA ATGGACAATC	900
TGTCTATATG GATACATGGG AATATTTAAA CGGCAAATGG TTAGTGAAAA AAGAACCTTT	960
CCGTAAATAT CCTTCTTACA TGGAATCATT CCAAGATAAT GCGCACGTGC TAAAAACAAC	1020
TTCTTTCCAA GCGGGCGTTT ACTATTATGC TGGGGCTTGG AAAAGCAATA CAAGCTCGTA	1080
CCGCGATGCA ACTGCTTGGT TAACAGGTCG TTATGCGACA GATCCTAGCT ACAATGCTAA	1140
ATTAAATAAT GTCATTACCG CATATACTT AACTCAATAT GATACACCAT CTTCTGGTGG	1200
AAATACTGGG GGCGGAACAG TTAATCCAGG AACAGGCGGC TCGAACAATc AATCAGGAAC	1260
GAACACGTAC TATACTGTAA AATCAGGaGA TACCTTGAAT AAAATTGCCG CsCAATATGG	1320
yGTGAGCGTT GCTAATTTAC GCTCATGGAA yGGCATCTCT GGCGATTTAA TTTTCGTTGG	1380
TCAAAA <sub>A</sub> CTy ATCGTGAAAA AAGGTGCTTC AGGTAACACT GGTGGCTCAG GCAGCGGTGG	1440
tTCTAACAAT AATCAATCAG GAACGAyAC GTACTAyAct tAAAtCAGGs GATACCTGA	1500
AyAAAAATTGC CGCCCAATAT GCGGTGAGyG TTGCTAATTT ACGCTCATGG AATGGCATyT	1560
CTGGCGATTT AATyTTCGyT GGTCAAAAAa TtATyGTGAA AAAAGGTACT TCAGGTAACA	1620
CCGGTGGCTC AAGCAATGGT GGTTC <sub>T</sub> AACA ATAATCAATC AGGAACGAAT ACGTACTACA	1680
CGATTAAATC GGGCGATACC TTGAACAAAA TTTCTGCACA ATTCGGTGTT AGTGTGGCTA	1740
ACTTACAAGC CTGGAATAAC ATCAGCGGTA GTTTGATTTT TGCTGGTCAA AAGATTATCG	1800
TGAAAAAAGG CGCCA <sub>A</sub> CTCA GGTTCACGA ATACGAACAA GCCTACGAAT AATGGTGGCG	1860

GTGCGACAAC ATCCTACACG ATTAAATCAG GTGATACGCT GAATAAAATT TCTGCACAGT	1920
TTGGCGTGAG TGTTGCTAAT CTACGTTTCAT GGAACGGGAT CAAAGGCGAT TTAATTTTTG	1980
CTGGTCAAAC AATCATCGTG AAAAAAGGCG CTTCTGCAGG TGGCAATGCT TCTTCAACAA	2040
ATAGTGCATC AGGCAAACGC CATAAGTTA AAAGCGGTGA TTCACCTTTGG GGCTTATCAA	2100
TGCAATACGG AATCAGCATC CAAAAAATCA AACAATTAAA TGGCTTAAGC GGGGATACAA	2160
TTTATATTGG TCAAACCTTA AAAGTTGGTT AATTTAAGAT TGAAAAAAAT AGCTATCTTT	2220
GGTAACATGA GTTTTATAAT AGAAAAAAC AATGAGAAGG AATAGTAGAA AGAACCATGT	2280
GATAGAGAGC GTATGGCTGG TGGAAATACG TACAGAAGCT TTTGAACTCG CCTTTAAGTT	2340
ACTTTTTTGA ACAAAACCAAG TAGGAAAAGT CGGTGACGAT CGTTAAGACG TTTGAGGTTA	2400
AGGTGACGGT GACTGACCGT TTCTTTGACG AATTTAGGTG GTACCACGTT GCATTTGTAT	2460
GTTACAGTCC TATAGGAATT TTTCTATAG GACTTTTTTT TATCACTAAG TGGCTAGTTT	2520
AAGTCAGAAT AGTGACAATG GCTGCAGGTG GCAAATTTTA AGAATGAAAA ATTTTATATT	2580
ACTAGGAGGA ATAACATGAG CTACAATCAC AAAGAGATTG AGAAAAAATG GCAAAAATAT	2640
TGGGCTAAGA ACAATTGTTT CAATACATTG GACGACCCAA ATAAAGAAAA ATTTTATGCA	2700
CTAGATATGT TTCCCTATCC ATCTGGACAA GGCTTACACG TAGGTCACCC GGAAGGCTAT	2760
ACAGCAACCG ATATTCTTTC ACGTATGAAA CGTGCGCAAG GCTATAATGT GTTGCAATCA	2820
ATGGGCTGGG ATGCGTTTGG TTTGCCAGCA GAGCAATATG CGTTAGATAC AGGAAATGAC	2880
CCAGCTGAAT TTAATAAGAA AAATATCGAA ACATTCCGTC GCCAAATTAA TTCACTAGGA	2940
TTCAGCTATG ATTGGAATCG TGAAATTAAT ACCACTGATC CTGAATATTA CAAATGGACA	3000
CAATGGATAT TTACAAAATT ATATGAAAAA GGGTTAGCTT ATGAAGCAGA AGTTGCGGTT	3060
AACTGGGTCC CTGAATTAGG AACTGTTATT TCAAATGAAG AAGTCATTGA TGGAAAAAGT	3120
GAACGTGGCG GTTATGATGT GGTTCGCCGA CCAATGCGTC AATGGATGCT GAAAATTACT	3180
GCTTATGCAG ATCGCTTATT AGAAGATTTA GAGCTTGTG ATTGGCCAGA GAGTATTTAA	3240
GATATGCAAC GAAATTGGAT TGGACGTTCT GAAGGAGCCA ATGTGACCTT TAAAGTCGCT	3300
GGCACAGAAG AAAGTTTCAC GGTGTTTACA ACCCGTCCTG ATACCTTGTT TGGTGAACCC	3360
TATACTGTTC TAGCTCCTGA ACTAGAACTA GTGAAAAAAA TTACGACACC TGAACAAACA	3420
GCAGCTGTAG AAGCATATAT TGAAGAAACC TCAAAAAAAT CTGATTTAAA TAGAACGGAT	3480
TTAGCAAAAG AAAAAACAGG TGTTTTTACA GGTGCGTATG CTATAAATCC AGTCAATGGC	3540
CAAGAAATTC CAATTTGGAT TGGCGATTAT GTTTTAGCAA GCTATGGCAC AGGCGCAATC	3600
ATGGCGGTCC CAGCACATGA TGAACGGGAT TACGAATTTG CGAAAACATT TGGCATTGAT	3660
ATCCTACCAG TAATCGCAGG TGGCGACATT ACAACAGAAG CCTATACAGG GGATGGACCG	3720
CATATCAATT CTGATTTCTT AAATGGATTA AACAAAGCAG AAGCCATCGC TAAAATGAAT	3780
GAGTGGCTAG AAGAAAATCA CGTAGGGAAA AAAGAAGTAT CTTATCGTTT ACGTGACTGG	3840

TTATTCTCTC GTCAACGCTA CTGGGGTGAA CCAATTCCTG TGATCCATTG GGAAGATGGA	3900
ACAACCACAA CGGTTCCCTGA ATCTGAGTTA CCTCTACGTT TACCAGTAAC ATCGGATATT	3960
CGCCCAAGTG GAACTGGGGA ATCGCCATTA GCAAACATTG ATGAATGGGT CAATGTCGTC	4020
GACCCTGAAA CTGGCATGAA GGGAAAACGT GAAACGAATA CTATGCCACA ATGGGCTGGA	4080
AGCTCTTGGT ATTACTTACG ATTCATTGAT CCTCATAATA AAAATGAAAT TGCTGATTTT	4140
GAAAAATTAA AACGTTGGTT ACCAGTTGAT ATCTATATTG GTGGTGCCGA ACATGCGGTG	4200
CTGCATTTAC TTTATGCTCG TTTTGGCAT AAATCTTAT ATGATATTGG TGTGGTTCCT	4260
ACCAAAGAAC CTTTCCAAAA ATTATACAAC CAAGGTATGA TTTTAGGAGA AAACAACGAA	4320
AAAATGTCTA AATCACGTGG CAATGTTGTA AATCCCGATG ATGTGGTGGC TAAATATGGT	4380
GCGGATACGT TACGTCTTTA TGAAATGTTT ATGGGCCCCAT TAGATGCTTC CATTGCTTGG	4440
AATGAAAATG GATTAGAAGG AAGTCGTAAA TTCTTAGATC GCGTTTGGCG TCTGATTGTT	4500
GATGAAGAAG GCAAAATGCG TGACCGAATT ACCACAATTA ATGATGGCCG TTAAACGAAA	4560
GTTTATCATC AAACGGTTAA AAAAGTGACA GAAGATATGG CAAACTTGCA CTTTAATACA	4620
GCGATTTCTC AATTAATGGT TTTTGTGAAT GAAGCCAATA AAGTGGATGC CTTACCTTAT	4680
GAATATGTGG AAGGATTTGT CCAATTACTT GCGCCAATCG CGCCACATAT TGGTGAAGAA	4740
CTATGGCAAA TTTTAGGTAA CGAGGAAAGT TTAACCTATG TCCCTTGGCC AACCTATGAT	4800
GAAACGGCCT TAGTAGAAGA TGAAGTGGAA GTAGTTTCC AAGTGAACGG AAAATTACGT	4860
GGCAAACAAA ATGTCGCTCG TGGGTTAAGC AAAGAAGAAT TAGAACAAAT TGCAATGAAC	4920
CATGAAGCTG TTAAAGAATT TATTGAAGGA AAAACAGTGC GCAAAGTGAT TGCTGTTCCA	4980
GATAAATTAG TAAATATTGT TGCAAATTAA GTTTACATGT TTTTATTAAA AAAAGAGCCG	5040
GACGTATCAA AACGTCAGGC TCTTTTTAAG GTTGTTTTGT TTTACCTTTT TGATTCATAT	5100
GTTCTTTACA TGTGATTGTT ATGACTGAAA GCAAGGCGAT GAAGAGAAAA AAGTAACCAA	5160
GCGGCGTACA GAAAAAGGCG GGTTCTTTTA AAAGTCCATC GCTAGATAAT GTTGATCCAA	5220
CTAAAATAGA GATCATAAAT AATAAAAGAC TAATCAAGCC AAGAAATAAG CTAATACGAT	5280
ATTTTTTCAA CTTCAATTCC TCCTTTAATG AACTATCACT AAGATAAATC AAAAAGGTTA	5340
ACTTTTTTTT AGGCAACTGT TAAAACTTAG ATAAGAAATC TATTTTCTAA GTTTGTCGTT	5400
GTTGGTAATT GGTTAAGATC ATACGTGAAA AGGGACCGAC AATCAGCAAT TGATAAGGAA	5460
GGGCAACAAT AAAGTTCAGG CCCCAAGCTT GGAAATACGC ACTTAGTGAC AACGTCTGGC	5520
CTTCATTAA TAAGCCAAAT AAAGACATAC AAGTCACCAT GCTGATAATC ATGAGACTCG	5580
AAATGGTCAG AATAAGTAAT AGCCGTTTTT CTTTATTAAA GGTTAAAGAA AAAGCGATTT	5640
TCTTAGCTAA GACACCAACG ATGAAAACAT CTAATAATAA AGCAACGATA AATCCAGGAA	5700
GCAAACCTTT TGATAAGGCG TACAATGAAA ATTCTCCATG CAATAATAAA TTATAGGTGC	5760
TCATTCCGAA AACCATCAGA AAGCACATTA ACGTAATAAA AAGCATTCCT TCTTTTTTAT	5820

TAGTAGGCAT TTTTTCTCC TTTCATTTAT TGAGAGTAAT GATTGTTCTA GGCAGAAAaG	5880
GACAAATTCT CTCGACTTTT ACTAGTGTAG CAAAAATAAA AGTTCTTCAT AAGTTAGAAT	5940
TTGCTAAGAA TTTGATTCAA ATAAAAAGGT GTACTAAGTC AGCAGTACAC CTTTTTATTC	6000
GTTATTTAGT TTATTTTTGT GTTATTTGTA ACGTCCTCTC CAAACCATTT TTGAGAAATT	6060
TTACTAAGGG TGCCATCTTT TCGTAACGTT TCAAAGGCAG TATTGATTTT TTGGACTAAT	6120
TGATTGTCTG ATTTGCGGAC GCCCACAGCA AAATCTTCAT TGTCATAGCC TACATGAGAA	6180
ATAGTATAGT TTTTAAATT ATCTTCGTGG GAAAGATAGT AGTTGGCGTA AACGCGATCG	6240
ATTAGGAGTC CGTCAATTCG ACCAGATTTT AAATCTAAGA AAGCTTCATT AAAGCCGTCA	6300
TATAAAATAG GTGTTTGGTC TTTAACAAAT TTTTCAAAA CGTCAGGCTG ACTTTCGAAG	6360
CCATCATAGC CAGAAGAGCC GTTTTGAACC CCTAAATTTT TGCCTTGCAT GTCGCTCGCT	6420
GTTGCAATGT TTTTTCTTT TAAAGAAACA AGTACTTGGT CGTTCGTCAT GTAAGGTTGT	6480
GTGAATTGAA CTTTTTCGGC CCGCTCGCTC GTTTTAGTGT AGCCGTTCCA AATAAGATCA	6540
ATGGTTTGAT TTTGTAATTC TGTTTCTTTC ATAGACCAAT CAATCGGTTG GAAGTCAACG	6600
GAAATGCCAT AAAGTTTAAA AACCGCTTTG GCTAAGTCGA CATCAAAGCC GACAATTTTG	6660
CCTGATTTAT CTTGAAAACC CATGGGCACA AAGGAGTCAT CTAAGCCAAT AATAATCCGT	6720
TTTTCTTCGT TAATCCGTGT CCATTGATCT TCGTTGCTTT TTCTTTTACC ACAACCTGCA	6780
AATAGGAGTA AACTACAACA GATAACCAGC AAGGCTAAAG AATATTTCTT TTTCATCATA	6840
AACTCCTTTC TATTGGACGG GGGCAACCGT CAAAAGATTG TCTGCAATTT TTTCAGCGAA	6900
AGCCATATCG TGGGTCACGA CAATCTGTGT CATGCCTTGT TTCCGTAAAT CTAAAATAAC	6960
TTCTTCCACT TGTTGGCGTA AAGCTGGATC TAGCGCGCTC GTTGGTTCAT CGTACCCAAG	7020
AACTTTTGGC TTCATTGCTA AGGCACGAGC AAGTGCAACT CGTTGTTTTT GTCCACCGGA	7080
AAGTTGGTAC GGATAGAGTT TTTCTTTACC AGCCAAGCCC AATTTTCCA GTAATCTAA	7140
AGCTTCTTGC TGACTTTTTG CTTTTCTTC TTTTAGTACT AATGTAGGTG CTAAGGTAAT	7200
GTTTTCTAAA ACTGATAAGT GAGGGAAAAG TTGGAAATCC TGGAAAACGA CCCCACCAC	7260
TTGATCGGCA TTGTCCATTT CaGCAGGATT GAAAGGAACA CCATCCAACA AAAGCTCACC	7320
AGAATCAATG GTTCTAAAC CAGCCAAACA ACGAAGTAAA GTGGTTTTTC CGCCGCCTGA	7380
AGGACCAACA ATCGTTAAAA TTTTACCATC TGGAATTCT AAATTTAAGT GATCAATAAT	7440
GGTCCGGTTG TCAAAATGTT TGGTAAATT TTTAATAGTT AACATGGTAG AGTCCTCTA	7500
TTTATAATAT TGATAGAATT TTTCAATTTT CTTAGAAGCA ACAGTCAAAA TTGCCGTAAA	7560
GAGTAAATAA ATCACTCCCA CTAAAGCTAA TGGCACTAAT GTAACATCGC GGCTCATCGC	7620
AATTTTACCA GCACGTAAAA GATCGCCTAA GCCAAGAACA TAAATAAGAG AAGTATCCTT	7680
GACCAAATTG ATAATTTTCAT TTCnTACGGA AGGCAGCACG ATTTTAACAA CTTGTGGCAA	7740
AATAATCCGC GTGACTGTTT GAAAACGTGT TAAACGTAAA ACCTTAGCAG CTTCATATTG	7800

ACCTTGGGGA ATTGCTTGAA TCCCCCGCG GAAAATTTCA GCAAAGTACG CAGCATAATT	7860
TAAATAAAT GCGAATAAGG CCGCATCGTA ACGTTCAAAG ACAATACCAA TTAATGGTAA	7920
GCCGTAAAAA ACAAAAATTA ATTGTAACAA TAGCGGAGTG CCGCGCATT AACC AATGTA	7980
GAGGTTAATT AAGTAACGTA AAGGTTTAAA CGAACTTTGT AAAATAAAGG CAATGATGAT	8040
CCCCAGTGA ATCGAGCCAA GGAGCGTAAA GAAAAAGAmT TTCAGTGTC TGCCAGTCCG	8100
CTAACTAAAG CAGGTAAAT TTCTAAAATA TAATCCATCA TTTTTCCTC CAACAATTTT	8160

## (2) INFORMATION FOR SEQ ID NO: 160:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6253 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

CAnCCCAAT TnTAGCTTG ATTCTGCAGC TTTTGATAAC aTTTGGTCAA TTTTGGTAA	60
TCCATTAAAT CAACTAACTG ACCTTGTTTT AACATGkkrS tAAAGTCTAA ATCCCAAAGC	120
ATTTCTTTAG TCGGCTGTTG CAATAAATAA TACATAGAAT GACTACTAAT AATTCAGTG	180
GTCGGTTCTT CAATTGTTTC ATAAAAAGTC ACAAATATT CTCGCTCAAT CACAGTAATA	240
ATTTCTTCTT CTGAAAGCTC ACAATCTTCT AAATTTTGAT AGAATATCAA TAATTTAATC	300
CGTTCATCTA TTTTATTCAG CGAancGACT AATTCAGGAA AAAGAACCGA CTGACAATTT	360
TCTACTAATT CTCTGCCTT ATTTTGTTCA TTTTATAAA AGTTGTATAA CTTTTCCGC	420
TGTTTTAACG AGTTTAGTAA GTTATCGGGT ATATTCCAGT AATATTGTTT CTTCATAGCA	480
TTCCCACTCT TCTTTTTTAT GTACAATTTT ACAAATTTGG TCTTAGCCAA TTTCTCTCT	540
AGTTATTCTA TCAAAAAGT GATGCTTTTT TACTGACaAT AATTACTCGT TTTACGCCCCA	600
AAAAGTGAAA ATCTAGCTGT TAAGTAATTA TTGGGGCTTG TTyCCCTAGG gCCGAaAAAA	660
CTTTTTCTT TTAAGTCC AGAAAAAGAC CTACACTAGA GATAGACAAA CGTTGGAGGG	720
ATTTAAATGG ATTTACACTT AACGAATAAA TTAGCATTA TTAATGGTTC AACCAAAGGA	780
ATTGGCAAAG CAATTGCGAT TGAAATGGCT CGTGAAGGGA CCGATGTCAT TATCAATGGG	840
CGCAATGAAG CCGAAGTAAC TAAAGTTGTT AAAGAAATAC AAACAATGTT TCCAGACACT	900
CATCCTCAAG CAGGGACTGC CGATATTTC ATTGAAAGTC AACGAGCTAC TTTGCTTGAA	960
AAATCCCTA AAGTCGACAT TTTAGTGAAC AATATGGGGA TTTTGGAGCC AATGGAATAC	1020
TGGGACATCG ATGACGCCAC TTGGGAAAAA TTTTACTG TGAACGTGTT GTCAGGCAAT	1080
GCATTAGCAA AAGCTTATCT ACCTAAAATG CTTGCACAAG ATTTTGGTCG CATTATTTTC	1140
ATCGTAGCG AAGAAGCGGT GATGCCTTCT GGCGaAATGC CCCAATATAG CATGACAAAA	1200
ACGATGAATC TTTCTTAGC TAAAAGTTTA TCCAATTAA CTGTTGGCAC ACATGTCACC	1260

GTTAACACGG TTATGCCTGG CTCAACCCTT ACCGAAGtGT AGAAAAAATG TTGGAAGATA	1320
TGTACGCTGA TTCAGACATT CCCAAAGAGG ATTGGGAAAA AGATTTCATG AAAAATCATC	1380
GTTCTCGTTC ACAAATCCAA CGGCTCATTC GACCAGAAGA AATTGGTCGT TTTGTTACCT	1440
TTGTGGcCAG CCCAGATTCT TCTTCTTTCT CAGGAGAAGC CTTAAGAATC GATGGCGGCT	1500
TAGTTCCAAC AATCTTCTAA AATAAACAAA AAAGACAGCC AGCAAAAATT CACATTTTTn	1560
GCTGGCTGTC nTnTTArGCA TATTAATTGA TTACTIONCCA AAAGTAAGTT ATATTATCT	1620
CGAATTCGAA wTAAAAAGAG GTGAAMCAAT GAACCAACAG CAAGAAGCTT TAAAAGCCTA	1680
TATCGGTTTA TTAAGAACCA GCCATCGACT AGAGCAACTT GCCAAGCAAG ATGTCACCTG	1740
TTATGACTTA AACATTACAG AATTTTCAGT ATTAGAGCTG TTACTIONATA AAGTCTCTCA	1800
GACCATCCAA AAAATTAAGG AGAAAATTTT AATCGCTAGC AGTAGCACCA CTTATGTTAT	1860
TGACCAATTA CATAAAAAAG GCTATGTAAC GCGCACCCCC AGTGAAAAAG ACCGACGCAT	1920
TACTTACGTT GAATTAACAG AAGCTGGAAA AACATTAATT GAAGAAATTT TCCCGACGCA	1980
TGCAAAGCGA ATTGCAGAAG CATTTGAACA ACTCTCTTCC GAAGAATTAA CACTTCTTCA	2040
AAAAACTTTA CGAAAAATAA CAAATGAaAC GAAATGAGGA AATAATGATG AAAAAAGAAG	2100
ATCAATTATT AGGAATCCAC CACGTTACAG CTATGACAAG TGATGCAGAA AAAAACTATC	2160
ACTTCTTTAC AGATGTTTTA GGGATGCGTT TAGTCAAAAA AACAGTTAAT CAAGATGATA	2220
TCTATACTTA CCATACCTAT TTTGCTGATG aTTTGGGTAC ACCAGGTACA ACCaTGACCT	2280
TTTTCGATTT TCCAATAAC CCTAAAGGAT TAAAAGGAAC CAATACAATT TCAAGAACAG	2340
GGTTCCGGGT TCCTTCAGAT GCAGCTTTGA CTTATTATGA AAATCGCTTC AaTGAATTTG	2400
CTGTCAAACA CACAGGCATT TCTGAAGAAT TCGGGAAAAA AGTCCTTCGC TTTTGGGATT	2460
TTGATGATCA AGCGTATCAA TTAATCTCTG ATGAATTAAA TCAGGGCGTT GCAGCGGGCA	2520
CCCCTTGGA AAAAGGACCT GTTCCAACAG AATTTGCGAT TTATGGCTTA GGACCTGTCTG	2580
AAATAGCTAT TTCCTATTTT CATGAATTCA AAGAAGCTTT TGAAGAAATT CTAGGCTTTC	2640
ACGTAGTGGC ACAAGAAGGC AATCGCTATT TACTAGAAGT TGGACAAGGT GGCAATGGTG	2700
CCCAAGTCGT TTTGGTAGAC GATGATACTA GCTCACAAGC GCAACAAGGA TATGGTGAAG	2760
TACATCACGT TGCATTCCGC CTAGCGGATC GTAAATCACT TGGGACTTGG CAAGCGCTCT	2820
TTGATCATTT AGGCTTACAA AACTCTGGCT ATGTCGATCG TTATTATTTT GAATCATTGT	2880
ATGTTTCGCAT TGGGCATATT TTAGTCGAAT TAGCCACCGA TGAACCAGGA TTTATGGGGG	2940
ATGAACCTTA CGAAACATTA GGAGAAAAGT TATCTCTTGC GCCATTTTTA GAAAACCGTC	3000
GTGAATATAT TGAGAGTGTG ATCAAGCCTT TCAATACAAA ACGAGCCTAA GGAGGAAAAA	3060
ACATGCATTC AaTTTTTAAA AAAGGACATC CTGAAGCACC TGTCTTTGTG CTACTIONCACG	3120
GTACAGGTGG TGATGAAACA TCTCTCCTAC CAATTGCCCA AGAACTAAAT AAACAAGCTA	3180
CTGTGCTAAG TATTCGTGGT GATGTTTCAG AAAATGGAAT GAATCGTTAT TTTAAGCGCC	3240

TAGCGGAAGG TCATTATGAC TTAGAAGATC TAGAAAAACG CGGCGAGGCG CTTCATAAGT	3300
TTATTCAACA AGCCGCTAAC GAGCATCAAT TTTCATTGGA TAAAATTATT TTTATTGGCT	3360
ATTCAAATGG GGCCAATATC GCTATTCAAT TATTGCTTAC TCATCCCGAC AGCTACCATC	3420
AAGCTGTCCT CTATCATCCC ATGTTTCCTG TTGAATTGAC CAATCAACCA GACTTGACCG	3480
ACACTTCTGT TTTATTATCT CTAGGAGAGC ATGACCCGAT TGTCCCACTT CCTGAGAGTA	3540
TGCGTGTGAT TCAACTATTT CAGAATCATG GAGCAACCGT ACAAGAGGTT TGGACACAAA	3600
GTCATCAATT AACTTATCAA GAAATTAAGG AAACACAAAC TTGGTTGGCA CATCTGTCCT	3660
CTTAACAAAA AAGTGGGTCC TAAGTCATTT CGACTTAGGA CCCACTTTTT CTATTTTAAG	3720
ATAATTCACT GTCTGAGCTT CTTTTTATTC TGCCATTGTA ACAACTATTT TACCAACTGC	3780
ATGGTGC GTT TCACTGAGTG CATGTGCATC ATAAATCCCT TGTCTAGAAA AAGGGAAAAC	3840
TTCACCAATA ATTGATTTAA CTTTTCCAGC TGCCATCAAA TCAGCAATTT TTTGTAATTG	3900
TTCTCCATTT GGTTGAAGCC AAATACTTTC AGCAGAAATA TTTTTTTCGG CTGCCAATTG	3960
TTTATCTTCA ATTCCCACAA TTGAAACAAG ACGACCGGTG TTTGGTTTTA AGACGGCAAA	4020
ACTATTTTTT TGAACCTCAC CACCCATTGT ATCAAAGACT AAATCAACAT CAGTTAATAC	4080
CTCTGCAAAA TTCGTTGTAT GATAATCAAT CACTTCATCT GCACCAATTT TTTTCAGTAA	4140
GGTATGATTT TTTGCGCTAC GGTcGTGATG ACATGTGCGC CCGCTTCTTT TGCTAGTTGA	4200
ATCGCATAAG TACCGACCCC ACCTGCACCC GCATGAATTA AAACAATTTT TCCTTCTTTA	4260
AGGTGACCAT GATCAAACAA TGCTTGCCAA GCGGTTAAAC CAGCCAACGG AACGGCAGCC	4320
GCTTCTTCAA AGCTAATTGT TTCAGGGATT TTtGCTAATA AATGATCATC CACAATCGTT	4380
ACTTCTGCAT AGGTACCAAA ACGAGTAGTT TCAGGACGAG AAAAACTTT ATCGcCAACT	4440
TGCCAATCCG TTACTTGACT CCCcACCTCT GTAATCACAC CAGCGACATC CCAACCAAGA	4500
ATAATCGGAA AGGACCAATC AAACATCTGT TTtAAATATC CTTCACGCAA TTTCCAATCA	4560
ATCGGATTAA TTGATGTCGC GTATTCTTTT ACCAACACTT GATGTTCCGA TAATTCTGGC	4620
AGTGTTACTT CTGCTTCTTC AAGTACCTCT TTACTTCCGT ACTGATTAAT CACAACAGCT	4680
TTCATAAACC AACTTCCTCT CATTTGTTTC ATTTTTCACA CTTTtagTAT ACGnCCtATT	4740
GCCTTCTTTT TATATAAGAA TATGCTTTAG TTAAATTGAG AGAAAGTAGC TGTTTTTAGT	4800
CTATTATTCA TAaCGTAAAG ATTCAATTGG ATCTAATTTT GgCTGtCTTC GAGCTGGTAA	4860
AGTTCTGCT AAAAAAGCaA TAAACATAAT AACTAAGATA ATCGTTAGTG AAGATGGCAA	4920
TGAAAATTGA ATTAATTTAA ATCCTGTTAA AGCCTTTAAA AAAGAATCCG TTGCCAAACG	4980
ATTAATAAAA TTTCCAACCC CAACAGCTCC TAAAATGCCT AATATTGAGC CAAAAAGCC	5040
AATTAAAGCC GCCTCaACAC TAAAAATTGT AAAGAcTTC CCATTGCTAA GACCCATGGC	5100
TTTCATTAAA CCAATTTCTC GCGTCTTTT TTGGACAGAC ATATACAGCG TGTTAATAAT	5160
GCCAAAGcTT GCAGCTAACA AGGCAATTGC ACCAAACATG GTTAAGACAC CAGTGATTGC	5220



905

ATTGATTATA TTACGAATCA TGCCAATTTT ATCTTCGACA GTAGTCGCTA AATAGCCAGC	5280
TTTGTCTAAA TCTTTTTTGA TGTCCCTAAT CTGCTCTGGC GTGCTGTCTT TTTTAACTTC	5340
AGCAATAATC ATCGCGTATT GATTTTTTTAA ATGTTCTGGA AGATCTGCTT GATTGATTGA	5400
GACAACTTTA TCAATCAATG CTTTATTCAT CAGCGAGAGC CCATTTTGAA TGACACTGGC	5460
ATTTCTAACG CCAACAATTT TTGCTTCAAT AACTTGCTCT TGTCCCTTCA AGGAACTTGA	5520
GATACCTAAT TGAAGTGTTC CCCCCACGGC AGCTTTACTA GATGTATAGC CGAGCGCTTT	5580
TACATATTCT GGTGACAAAT TAATTTTCGAA GTCCTGGCTC GTTTCGAAA CTTTACGGCC	5640
AGCCGCTAAA TCAATGGTCA TCTCATCTAG CGCAGAAGTT GCTGAAAATA CATATTTATG	5700
CTTATCTGCT CCTTTTATAT AATCGATAGC AACGGATTTT ATTGGTTCCA CTGATGTAAC	5760
ATCCGAGATT TTTTAACTCT TCTCAATATC TTTTTCAGCA AGCATACTTT GTTGTGAAT	5820
CGTACTGGTT TTTTCTCTG GATTATATTT ACTTGGTTCT GTTCCGTTAC CAACATTCAT	5880
CTCCATTTTT GGCTGAATAA ATAAGTATT TGCACCACCG AACTGCCGA CTTGTTTGTC	5940
AATATAATCA TTTACACCAA TATTGACTCC TGTAGTTAAA GCTATGGTAA ATGCTCCAAT	6000
AAAGATTGCA ATAATGGTCA ATACGGTTCG ACCTTTATTA CGCATTAAAT TGGTACTAGC	6060
TGATTTTAAA ATATCTCTAA ATTTCAATTA CTCATCTCCT CCTACAATCA AGcCGTCTCG	6120
AACATGAACT TGCCGATCAC AACGTGCTGC TAAGTCTGGA TCATGTGTCA CAATAATtAa	6180
GGTAATcCTT tATTTTTATT TAAAyCAAaT AACAAATCTT CAATTTTCTT CCCAGTGGCA	6240
GAATCTAAAT TCC	6253

## (2) INFORMATION FOR SEQ ID NO: 161:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3312 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

TGTGGCAACT TAAAAAAGAA CAAAATGTTT CCATTTATCG CAGTATTATG GAGATGATTG	60
TTCAAAAGAT TCAAACAGGC GAATTATTAC CAGGCGAAAA AATCCCTTCG GAACGTAAAC	120
TGGCAGAGTA CTTGTCTGnC AATCGTTCAA CCGTTGTTCA TGCATTAGAC GAACTGGTGG	180
ACTTAGGCTG GATTGTGCGA AAAAGAGGGA GTGGAnGTnT GTCAACGAAG GAAAATGGGG	240
GGTAATGATG ACGCCACGAA CGGATTGGCA TCGTTATTTA GAACAAAATG TGTTCAAACA	300
AGTTCATCCC TTGGTGACAG AGATTGAGTT ACGTGCTAAA CAAAATTTGC CGACTGACTT	360
AGATATGTAT ACAGGCGAGT TGCCGTTAGA CTTAATTCCC AGTTTCGATT TTCCAGCATT	420
AAATTGAAA CAATTTTTTAA AAGAGGAAGC ACAAGATGAA TTAGGTTATT TGCCGTTAAG	480
AAAAGAAATC CAAGCCATTA CAGCAACGGA ATATCAGCTT CATTTACCAA GTGAAAGTTT	540
ATTGCTTACT TCAGGTGCGC AACAAAGCACT TTTTTTAGTT TTACAAGTTT TATTGAGACA	600

AGGAGATAGT GTTGCCGTAG AAGATCCATC TTTCTTCTAT GCCTTACCAA TTTTCCAAGC	660
TGCTGGTGTG CGCTTATTTG GTGTGCCCAT GACAGAGGAA GGGATTGATC TTGAAGCGCT	720
AGAACAAACA ATTCGACAAC ATCGCATCAA AATGGTTATG GTTAATCCTT CCTTTCAAAA	780
TCCGACTGGG ACAGTGATGC CTTTGAGAAA AAGAGAACAG CTTGTGAAAG TCTGTCAAAC	840
GTATCAAGTT CCTATTTTAG AAGATGATGT TTTTGGGCAG TTAAGTTTTA TCCCGAAAAC	900
AGAAATTCCT CCTTTAAAAA AATTAGATCC AGATAATGTC TTGTATATCG GTTCGTTGTC	960
AAAAATTTTA GGTTGACCA CGAAAATTGG TTGGTTAAGT GCGCCCGCCT CCGTGACGAA	1020
ACAAATTGCA GAAGCCAGAA AAATGATGGA TGTTTCGTTG AGTATTTTTC CACAGATGCT	1080
CGCTAAAATG GCGATTGAAG ATCCTTCTTT CTCTGAAAAA ATTACGTTAT TAAATAAACA	1140
AGTGGAAACA CGAGCAACAG CTGTATATCA AGTATTTAAA TCGTTGTCAG AATGGGAAGT	1200
TTCCGCAGTA AAAGGCGGCT TTTATTTATG GGCACATTGG CGTCAAGGGG CACTTAAGCC	1260
AGAAGATTGG CAAGTGTTTT TACGAGAAGG TGTTTTGGTA GCACCAAGTG TAGCTTTTAG	1320
TGAAAAACGT GGAAGTATTC GCTTGAAC TGTTCCGAATT TCACCAGAAG AAATGCCTCT	1380
TTTTTGAG CGAATGGTAC GGATTACAG ACAGCTTTCT GAAAATAGAC AAACAAAAT	1440
AGAACAGTAA GTAAGATTTT TCTTATAAAA ATCATTGCT GAGAGAAAAC GCTACAACCT	1500
CCCTTTTCTA AACGAATATC CAATGCTATA CTTAAAAAGT AAAACAAAAA AAACAAATGA	1560
GGTGTTTTCC ATGTCTTGGG AACAAAGTTT CCAACAATGG TTAGATGAAG AAAATATTCC	1620
AGAAAATTTA AAAATGAAT TAAAGGACTT AAATACAGAC CCTGAAAAAT GTGAAGATGC	1680
ATTTTACGCA CCATTAGAAT TTGGGACTGC CGGGATGCGT GGGATTTTAG GCGCCGGTAT	1740
TAATCGGATG AATATCTTTA CGGTTCGTCA AGCAACAGAA GGACTTGCAC GTTTTATGGA	1800
TACGCAAGAT CCTGAAACAA AACGTCGCGG TGTAGCTATT GCCTATGATT CACGTCATAT	1860
GTCTCCAGAA TTTGCAATGG AAGCAGCTAA AACATTAGCT AAACATGATA TTCCTTCTTT	1920
TGTGTTTGAA AGCTTACGAC CAACACCAGA ATTATCTTTT GCGGTCCGCT ACTTTAAAGC	1980
TTTTGCGGGA ATCATGATTA CGGCTTCTCA CAATCCAGCT GCCTATAATG GCTACAAAGT	2040
CTATGGTGAA GACGGTGGTC AAATGCCGCC AGCTGATGCA GATGCATTAA CTAAATATGT	2100
TCGTAGTATC GAAAATCCTT TGAAAATTGA TGTCTTATCA GATGAAGAAG TTGCCCACAG	2160
CGGTTTAATT AACATTGTTG GTGAAGAAGT CGATAACGCG TATTTGAAAG AAATTAAAAC	2220
AGTTACGATT AATCAAGAAT TAATTAACGA AATGGGTAAA GAATTAAAAT TAGTTTACAC	2280
TCCCTTACAT GGAAGTGGGA AAATGTTAGG AGAAAAGGCA TTAACAACAG CCGGTTTTGA	2340
AAAATTTGTT TTAGTACCTG AGCAAGCCGT TGCTGATCCT GACTTTACTA CAGTTAAGTC	2400
ACCAAACCTT GAAGAACATT CAGCTTTTGA ATACGCGATT CGTTTAGGGG AAAAAGAAGG	2460
CGCTGATTTA TTGATTGCAA CAGATCCAGA TGCTGATCGT TTAGGTGCTG CTGTGCGGAT	2520
GCCAAACGGA GACTACCAAG TATTAACAGG AAATCAATTA GGTTCATTA TGATTCATTA	2580

TATTTTAGAA GCCCATCAAC AAGCAGGCAC GTTACCTCAA AATGCTGCGG TTCTAAAATC	2640
AATCGTTTCT AGTGAACCTG CTACAGCGAT TGCTGAGAAA TGTAATACGA AGATGTTTAA	2700
TGTTTTAACA GGTTCCTAAAT TTATTGCTGA AAAAATTCAA CAATACGAAG AAGACCACAG	2760
TCAAACGTTT ATGTTTGGTT TTGAAGAAAG TTATGGTTAC TTAGTGAAAC CTTTGTTCG	2820
TGATAAGAT GCAATTCAAG CGTTAGTTTT ATTAGCAGAA GTAGCGGCTT TCTATAAAAA	2880
ACAAGGAAAA ACATTATACG ATGGTTTACA AGATATCTTT GAAGAATTTG GTTACTTTGA	2940
AGAAAAAACT ATTTCCGTAA CAATGAGCGG AATTGAAGGC AGCGGTAAAA TCAAAGCTTT	3000
AATGGCAAAA TGTCGTGAAC AAGCGCCAAC TGAGTTTGCC GGGATTCAAG TTGCTCAAAC	3060
GGAAGATTTT AAAGAATTAA CACGCACTTT TGCAGATGGT CAAACAGrAC AATTACAAAC	3120
GCCTCCTTCT GATGTGTTGA AATATCATT AGAAGATGGC AGTTGGATTG CTATCCGACC	3180
ATCAGGAACA GAACCAAAAA TTAAATTCTA TCTAGCGACA AAAGCAACAA GTAGTTCAGA	3240
AGCTTCTGAA AAAATTGCTG CGTTTGAAGC nGTCGTTAAT GAATTAACAA AATAAAATAA	3300
ACGAAACAAG CT	3312

## (2) INFORMATION FOR SEQ ID NO: 162:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14286 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

TATCTTCGTG AAATTCAAGA AGACTTAGCA ACCCTTCCTG AAAATACCTT AGAGATTCAG	60
ATCCATCACG GAAAAATTC ATTACATACG ACTTCCAATG ATTTCTTTC CCACTATTTT	120
ktAtTATTTA AtTatTaCTG kGtCCaTyCT ACTGACTTCa TyCTTTTTTCm AGcAATTATC	180
AAAAAAGrAA ACmATTCTGT CTGGAAAGTG AGTCmAGrAA CTAATTTTAG CTCATCTTAC	240
ATTTACAACg GTTAAaTAAC ATTAATAAAC TATTTGGTTT GTATGGAATT TCCGTCCATT	300
TTTCCCCCTC CGCGAGTAAA GTGATTACAG GTTCTGAATT TCAAGTCCTA TATGCTPTTC	360
TTGATGTCTA CTGGACCATC TTTTAAATA CCTTAGCCTT ACCACGTTAT ACCTCTACTT	420
ATTTTGAAGA TATACTGACG CAGCATATTA AACCAGAAAT TCTTGAACGC TTGGAACACG	480
CTTCTCTACT AAAATTAAAT TTACTTTTAC ACATCGCAA ATATCGCTTT CCATATACTT	540
GTGAAGATGA TCTCCAAAA GAATTGAAG AACATCCTTA TTTAAGGCTT TTCTCAGACT	600
CAAAACAAAC TATTTTTTGT TCCAATCAA ATTTTTCAGA AGATCAACAG CGGGTTGTAG	660
ATATCATTGC ACCCATTGCC ATTGATAATT TTGATTACAGC AAAACGAGCC AAAGATTTAC	720
TGGCTGCTCT AAATGaAATT TCTTGCCAG AATATTCTA CGTTGATTCT TTATTAACCA	780
ATTTTTGCGA aGTTTTTCAA ATTCCAAAA TGCGTGAATT TGATCGTCAT TACTTCTTAC	840

TTATTTTATT TAAAAATATT ATCTATAACC AATTATTTAT TTCACAGCAA CCAAATACCG	900
CAATTCTAGA TTTTtagTC AATAAAAATA GTACTTATCA AAAAGATTTA CAGAAAAAAA	960
TCTATGCTTT TTATCAAAAC TTTTCAGCGC ATTCCCCTTA TCAGTTAAGA TTTAAGGATC	1020
AGGAGAACCC AATTTGGACA TGTGAATTGT TGTTACGAAT TTATAAGCAC TATAGTCCCC	1080
CCAAAAAAT TAAAATTGGT GTTGCATATT CTCGCGATTT TTACATTGCT AGCTTTATTA	1140
TGATGAAAT CAAGCAAACA TTTAGTGAGG ATATTATTTT TGAACGAACC GATTTTCTG	1200
GTTGTGATAT TGTGATTACT GACTATCCCT TATTTGATTT ACCAAAAGAA ATAGACCGAA	1260
TTTACATTTT AGAAGAAGAA TTAACCCGAG AAGATTGGGA ACTAATTTTC AGTAAATAA	1320
GTATCGCCAT TTTTGATTG CAATATTAAA ACAACATCA AGCGCCAACC TGTCTACCA	1380
GCCATAATGA AAACAAGCGA CCTACAAACC ACTTCTGCCA GTTTTAGTCT GACAGAAAT	1440
AGTTTGTAGG TCGCTTGAT ATTACTGTTT TTTTCTATGT TACCTTCAA GACAATCTAG	1500
TAACGTACTT CAGTCCTTAA AGTGCTACTG TTCCACCGCT GATCAGTTTT TAAGCCTGAG	1560
ACGATGATCC AAAGTGTTTT TTGTCTCAGG CTCTTACATT TTTCCAACA AAGGTATTGc	1620
TaGGCGCTAC TTGAAGCTGA AAAGTAGTTT CTCTATCGAG ATTCTCATC AGTAGACTCA	1680
AATTTAAAC TTTTTCTAT AAAAGTGaCT TCTTTGCGT ATGTTTTTAC AAAAGAAGAA	1740
TAGTCTTTGG GATCTTTGAC CTTTCCAAT TGGTCAACAG AGACTTCTAA TTCAAATCCT	1800
GTCCCATCTT TAAAGAGATA ATATTTTGAG ATGTCTGTAT ATCCTTGATT CCGTTTTAAA	1860
ATATAGCCGG TTTGTACATC AGCAATTTTA TCAATTGGGT ATGCTTTGAT AACTCTTTA	1920
CTTGATCCA AATAATCAGA AAGACTGTTG ACAAGATTGC CAGAAAGGTA CATACTGCC	1980
GATGTATAAT AGTCTCTTTT TTCATAAAAA CTGATATTGC CTAGATATTT TGTATCTGGA	2040
TACTCCTCTG TAGAAGGTAA AATTA AAAAG TTCTTTCCGC CACTGGTTTC AATATCGGCA	2100
TCGCCGATAA TTTCTATTTG GTAGCCTTGA TTAAATAAAA GTTTATTTTC TTCTATTGTG	2160
ACTTTTAGTG GTTTTCTAA ATCAACAGGT CCAGTTGTGG TAAGTCCAGC TTCTTTTAAA	2220
GGATAACAT CTCTGATGA AGCGGAGTAA TCCAATAAT CATCTTTCTC CGCAAGATCA	2280
ATATATACTA TAAAGCCAAT TGCTACAGCA ATGCCTATCC AAATAATTTT TTTGTTCTTT	2340
TTTTGTTTTT TCTTTTTTTT CGTCTGTGGT TTTACCACTT TTTTATTGGG GTTAGGAAGT	2400
TGCTTTGCTG TATCTGTTCT TGTGACAGGA AAAACGACC ATTTTGGCTT TTGAAAAGC	2460
GCATCATATG kATCATAAAA ATTCTTTTGT TTAGCCATTT TTTCTCTCCA CTAATTTTTT	2520
TGATAAGGCA ATTTCTCAG ATGCTTCGGA GACTAATCTT TCATAATCAT CTGTGATACT	2580
TTCCGACAAC ACTCGTATGG TACTTAAGAT ACCATTGGTT GCAAGTGTTA TTTCwTCTGT	2640
TGACAATTTA TCTCTTTTAA TGCTTAAAAA CCGTTcTGAT GCTCGTAAAA TCTCTGGTAC	2700
CTTATGATAT AAAAAATCAC TAAAAAGAGT AATTGcTTCA GGATACTTCA CTAAGTGTTT	2760
AAAGATGCCT TcAGCGGAAG CAACACCCAG TTTTACTTGT TCAATTTTTT TTAAATTTcG	2820

TGACTTATTc	GTTAGTTCGT	TTAAATAAAT	AATTTGTTTT	TTTAATTCGT	CCATAATTCT	2880
TTGGTACATC	TCTAGGTCAC	TGTCCGTTAG	ATGGTGcTC	GTTTGATATT	TTTTTAGATT	2940
TTTTTTATAa	CGGAAATACG	TACGACCATT	TCTATAAGTA	ACCCAACCAA	TGTACCCAGT	3000
CCCCACTAAC	AAAACCATTA	TCCCATTCAA	ATCATGTACT	CCTTTTTAAA	TTAGTATATA	3060
TTACTGAATA	ACACACTGTC	CAGCTTAAAA	TTGACAAATA	AGTATGAAAT	GTTACTCCCC	3120
CAATTACTTG	TGCACTTTTA	TTGTATCATA	TAGGAAAATT	GTTCTGTTCC	GCATTTGAC	3180
AAATTGTCTA	AATGTCTAAG	GGAATCTTTT	TTTGTTTAAT	ATGTAGCAAA	AAACACTTGC	3240
AAGAATTGAT	TACTTCTCAC	AAGCGTTTTT	TATTGTCTAA	AACAACATT	TACATTCAT	3300
TTTATTAAAT	AGTGAATAAA	TAAATTACTG	CAGCTAATGA	AGCaCCTAAG	ATAGGTCCCA	3360
CAATTGGcAC	CCAAGAATAA	GCCCAATCGG	AATCACCTTT	TGTTTTCACA	GGAATAATTT	3420
GGTGCGCTAA	ACGAGGTCCT	AAATCACGTG	CAGGGTTAAT	TGCATAACCT	GTAGGTCCGC	3480
CTAATGATAG	ACCTAGTCCT	AAGACCAAAA	TCCCAACAAG	CATTGGGTTA	ATGCCAGCCG	3540
CTAAGTCGTT	TTGAGAAAAG	GCTAGTAAAC	CAAAAACCAA	GACAAATGTG	CCGATAATTT	3600
CTGTATAAAC	ATTGGCTGGA	TAGTTACGGA	CAGCTGGACC	AGTGGCGAAG	GTTCCATAAA	3660
TAGCACCTTT	GTCTTCTGTA	ATGTTCCAAT	GGGGTAAATA	AGCTAACCAG	ACAACTAAAC	3720
CACCAATAAA	ACCGCCTAGG	ACTTGAGCTA	CGATATAAGG	TAAAACCATG	CCCCATTCAA	3780
AGTTTCCAGT	AATTGCCATC	GCAACTGTTA	CAGCAGGATT	TAAATGCGCT	GGACTCATAT	3840
ACCCAGACAT	GTAACAGCT	AATGTAACAG	CAGCGCCCCA	ACCTAAAGCA	ATAACGACCC	3900
AACCAGAAGC	AAAGGCTTTG	CTTTTCTTCA	AGTTAACTGC	GGCACAGACG	CCATCCCCTA	3960
GTAAAACTAA	AATCATCGTT	CCGAAAAATT	CACCGAATAA	TTGTGTCATC	ATCGAAGTTC	4020
CCATTTTTTC	TTCACCCTAC	TTTCAAATTT	TTTAAATCTG	ATTCTTCAAT	TACTTTTTCT	4080
AATGTTTCAA	TGTGTCGTGC	TCTTTCTTCC	GCTGTCCACT	GATAATGCTG	TGCCATTTCT	4140
TCAATGACTC	CCGCTTTTAC	TTGGTCCAAA	CGATCACGCA	TAAATAATAA	GTGGTTGGTC	4200
CGTCGCAATA	AAAAGTCGAC	GGGTGTTAAA	GCCATTCTTT	CTTCCATCGC	ATAGTTTAAA	4260
GAGACAGTTT	CTGTTAAGGT	TAAACCAGGA	ATTACTTTGG	CATCATCAAT	CATTTCAAAA	4320
ACAGTTGGTA	CGTTGGAGCC	ATATAAATGA	GCTAAGTATA	GTGCATCCTT	TTCACTTAAC	4380
CCTTTAGACA	CGCCATGTTT	TGCCAACGCT	TCTAATTCTT	CGTCTACCGT	TGCAGGGTTC	4440
AATTTTCCGC	CTGAAACTGG	ATAGTTTTTA	GAATCAATTA	AGGTGAATTC	TTTCTGGTAT	4500
TCTTCTGCTA	AGATGGTTTG	AATCATTTTC	ATCGCACCTT	CTGCCATTTT	ACGATAATCT	4560
GTTAATTTCC	CACCAGCCAA	AGTGAGCAAG	CCATCCGCTG	ATCTTTCCAA	TGAGCTGCCT	4620
CGAGAGACAG	CTGAAGGATT	GACTTTATTT	TCCACCAAAAC	TGTCTTCCAG	ATGGTTTAGG	4680
ACTTCTTCAA	TTTCTCGTTT	ATCTGTTTGA	TTTTTTTGGT	ATCTGTCAAC	GACATCAATG	4740
ACTTCATCAA	TACTTTTATC	TGAAAGTTTG	CCATTGTTTC	CCCCATTATA	ATCTGAGCCA	4800

CCGTTAGCAG	AAATTAAAGG	TCGTAAGCCA	GCCCAGCTGG	CTTCAATATC	ATCAATTGTC	4860
AATTGTACCT	CTGGATAACG	ATTATTGACT	ACTTCCAATA	GGTAATCCAC	ATCACTTTGT	4920
TCTACTGTTG	GATGTTGAAA	GTCTCCATGG	TAATCTGTAT	CCGTCGTCCC	AAAATATGTT	4980
TTTTCTTCTC	TTGGAACAAC	GAAGACCATG	CGGCCATCTT	GTTTACCTGT	ATCAAAATAC	5040
GTTGGTTGAG	GAACATTTAA	ACGGCTTTTG	TCGACAACTA	AATGAACCCC	TTTGGTTGGA	5100
CGCATTTGTG	GTGTAAATGA	ATCGTTTTTG	TCTAAGCCTC	TTAATTTATC	TGACCAAGGA	5160
CCAGTAGTAT	TAATCACAAC	ATCTGCGTGA	ATCTCAAAC	GTTTCGTCAGT	TAGTAGGTCC	5220
TTTGCTTGAA	TGCCAGTAAT	TTTCCCTTCA	TCATTGTATA	AAAAGCCCAC	CGCTTGAACC	5280
CGACTCACAG	CATGCGCCCC	GTCAGCTACT	GCTTGTTTAA	TATTTTCAAT	AACTAAACGA	5340
GCATCATTAT	TTCGGTAATC	TAAATAAACA	CCGCCACCTT	GTAAACCTTC	ACTTTTTAGT	5400
TGTGGCTCAC	GTGCCAAAAC	CTCGTCTTTT	GTTAACAGAT	AGTTGGCGTA	CTTGGTACCT	5460
GTTACATCAG	CTAACTGATC	ATATAAATCC	ATTGCTATTT	TGACAGAAAA	TAGATTAAAT	5520
GTTGCATTTG	GCTCATCATA	AATCGGCAAT	AACATGGGAT	CCGCTTTAGG	GATGTGTGGC	5580
GCAATGGATT	GAACGACGGC	TCGTTCTTTC	ACTGTATCTG	CTACCACCTC	CACGTCAAAC	5640
GTTTTCAAAT	AGCGGATACC	ACCATGAACT	AATTTGGTTG	AtCGTGAAGa	GGTGCCCTCT	5700
GCGAAATCCT	GCATTTCAAT	TAACCCTGTC	TTCATTTTAG	CGGCACTGGC	TTGCAGGGCA	5760
ACCCCTGCGC	CAGTAATGCC	ACCACCAATA	ATTAAGACAT	CTAAGTGTTG	TGTTTTTAAA	5820
GCGTCGATTG	ATTGTCGGCG	TGTTTCTATT	GAGAAAGTCA	TAGAAAATTC	CCCTTTTCCA	5880
TTTACTTATT	TTTACGTTTG	AATTGTTGGG	TAGCTGCTAC	TGCTTGTTGC	CAACCTTCAT	5940
ATAAATCTTC	TCGTTCTTCT	TCCGCCATAA	TTGGTTCAAA	TTGTTGTCCT	TCTTCTTGGA	6000
ATGCTTTGAT	TTCTTCTAAG	TCTTTCCAGA	AGCCAACCGC	TAAGCCCGCC	AAGAAAGcTG	6060
CGCCTAAGGC	CGTTGTTTCT	AAGTTATGGG	CTCTTTGAAC	AGCAGTGTTT	AAAATATCTG	6120
CTTGGAATTG	CATCAAGAAA	TCATTGTTTG	CTGCTCCACC	GTCTACTTTT	AAGACGGGAA	6180
TATCTATGCC	TGTATCTTCT	TTCATCGTGT	CAATAATGTC	ACGAACCTGA	TACGCAACTG	6240
CTTGAAGTGT	CGCTTTGACA	AAATCTTCTC	TCGTGGTACC	GCGTGTTAAT	CCAAAGACAG	6300
CCCCACGCGC	TTGTGAATCC	CAATATGGTG	CGCCCAAACC	AGTGAATGCC	GGAACCACAT	6360
AAACTTCATT	GTGCCCTGTT	GACGCTTTGG	CTACTGCTTC	AGATTGAGCA	GCCGTTTGCA	6420
ACATTTTTAA	GCCATCTCGT	AACCATTGAA	TCGCTGAACC	AGCCACGAAG	ATACTTCCTT	6480
CCAAGGCATA	ATATACTTTG	CCGTTAATGC	CATAGCCAAT	AGTGGAAGC	AGATTATTTT	6540
TAGAAAGTTG	AGGTTCTTCG	CCAGTGTTCA	TGACAATAAA	TGAGCCTGTT	CCATAGGTAT	6600
TTTTAACCAT	CCCAGGTTCA	AAGGCCATTT	GACCAAATAA	GGCGGCTTGT	TGGTCACCAG	6660
CCATCCCAGC	GATAGGAACT	TACTTCCGT	AAAAATGGTA	ATTTTTTGTT	AATCCATACA	6720
CTTCAGAGTT	CGAAACGACT	TTCGGTAACA	TACTCGTGG	GATATTTAAA	AGATCCAAAA	6780

TTTCTTGATC CCAATCTAAA TCATGGATAT TAAACAACAT CGTCCGACTT GCATTTGAAT	6840
AATCAGTGAC ATGAGTATCT CCCGTTAATT TCCAAACGAG CCAAGTGTCA ATTGTCCCAA	6900
ACATCAATTC TCCATTTTCT GCACGTTCTT GCGCACCTTC TACGTGATCT AAGATCCAAC	6960
GAACCTTCGT TGCTGAGAAA TAAGCATCAA TGATTAACCC AGTTTTTTCA TGAATCATTT	7020
CGCTGTAGCC ATCTTCTTTT AGTTGATCAG CAATAGGTGT CGTTTGGCGT GATTGCCAAA	7080
CAATTGCGTT ATAAATGGGT AACCCAGTTG CTTTGTCCCA AACACGGTT GTTCCCCTT	7140
GGTTTGTAAT CCCAATTCCG GCAATATCTG TAGGTTTCAC TCCTGACTCA ATTAAAGAAC	7200
CTGCAATAAC GGATTGAACA GAATTCCTTTT TTTCATTGGC ATTATGTTCC ACCCAGCCTG	7260
CATTAGGAAA ATATTGAGTA AATTCCTTTT GTGAACCTACC AATTTTGTTT CCTTTTTTAT	7320
CAAAAATAAT TGCACGAGAA CTAGTCGTTT CTTGATCAAT CGCCATAATG TATTTTCTT	7380
CTGCCATGAT AATTCCTCCT AAACGTTTTT GAAATCGTTT TCTTTATGTC CCTATCTTAC	7440
AAATAAAAGA ATAACTAAAC ATGTCAAAAA CTTTTTTTAT TTTAAAAAAA TTGAAAGGAT	7500
TTCTACTAAA ACCAACAAAG ACAACGCTTT CAATAATACT TTTATTATCT CTAATTTTTC	7560
AATAAAAAGA AACATTTTTG TTTAAAAAG CTAATTTTTC GACTTTTAAT TGTAAAAAGG	7620
CTGGTGCGAG GATCAGACAA TTTCCGTCAC ACCAACCTTT TGTTTATTTT TCCCGCAAGG	7680
TGCGTATTTT CTTTTTTAAT TGTGTCACAT CATAGGTCGA ACTTAATTCT GACAAAACAT	7740
AGACTTCCGG TGTTCTTTGA TAGCTCGTTT GCTGATAATT GGTCACAATT AAATCATATG	7800
TTTGTTTCAAG CTGATACCTC TCCACCAGCG TGCCATTCAT ATTTTAAAGA CTAAGCATT	7860
ATACTTGTGT TAACGCCGCT TGATAAAGAG CATCCATTGC TAAATGAATT CCAATTTT	7920
TTTCACCAAT CATTTTAAAT TCAATTAATG CCAACACGCT TAAATATTG ACCAGAGAAA	7980
GTTTCATGCA ACTATCTCCT GGTGATACG TTTTCATCAA ACAAGCTAAA GCTTTTTCTA	8040
ATAACGTTGT TGAAAAAGAA GCTAATTTCC GACTAGATAA TTGCTGTTCT TTTTCCGAGA	8100
TATGGGCCCC ATCAAAGACT TCCAATTCTC CTTTAAAAAA ATACAGTTT CCATGAATTT	8160
GTGAAAAATA GTAAGCCATG GTTTTCTCAA GAGTAGGTTT AAATTTATGG GGCCGATAAT	8220
ACAAGACCAC ATTTTCCAAG ATAAACGTAT CAGCTAAAGA AACAGCGGAA CGACGTCCCC	8280
GAATTAAGGA ATATTCATCA AAAGCTGATT CTGGTAAGAC CGACATTGTG ATTAGAAAA	8340
CAAAATGAAT CATGCTTTCT CCTTCGTCTA ATTCTAATGA ATAGCGACTA GAAAAACGGA	8400
AAAACAATTG TCGGACTTGC TTAAATAACG GATCTTCTTT AAAATCAATC ATTTTCTTTC	8460
TCAATTGTTG ATATTTTTTT GAAGTGACCG CTAGGCGCTT TTTGTAATA CTCAACCAA	8520
GGCTAACTTT TAACGCATTA TGTTTATTAA AAGTAAACC TAAGCCAGTC TCAATCCCTT	8580
CAATTATCCG TTTATTTAAC GGCGACATCG TTTTTTCTG GTGTATTTCA TAGGGAGTAA	8640
TATACCAATA CAACTGAAAG TAAAAGTAAC GAATCTGCAA TTCTTCTCCC TGTAGTTGTC	8700
CATTTTTTAT TTGCAAATCA AAGACGGCTA ACAATTGATT GAGTTCCTTA ATCTTTCGAA	8760

ATAAGAAGA CTCACTAAtC aTAAATTTAG TCGtTAATTG AACAAATAGAA AACTCTTTAT	8820
TACGATATAA ATAATCtAAA AGTTGAAATT TAAGAGACTC TTTTACAAAT GCTTCAATTA	8880
AATGATTCAA CGAAAAA TCCGATAAGG TGATTGATAG CTGTTGGCCA TCATAAGTAA	8940
GCTGACAATC CTGTCCAAAA GGTTCAAAA AATAATGTAA ATCCTCTAAG TCTTTTTCAA	9000
AGGAGGcTTT TGTGACCGCT AAAAATGCTG tTAAATCAGA GACTGAGCAG GTGCCGCCTG	9060
CTAAGATGGT TTGTTGTACC GTTAATAATT GTCGCATCGC TTTCTTTTct AATAATGTCT	9120
CAACTTTCAT CGTTCTCTCT CCAATAAGCC GAAGCACCCA TCAAATTTGT AATATCATAA	9180
ATATTGGTTA AGGAAGGTTG ATAAAAATAA TCTTTTTGTA ATAGGTCTGT CAGTGTCTGA	9240
CCTGTTTGA TACTTAATGC TAAAGTATTT ATTTTTTCTA AACAAATTGTT TTTTGAACAA	9300
AGCTGTGCAC CAAGGACCCG TTGCGTCACT TTATCATAGA TTAATTTGCC TAAAATTTCT	9360
GTCCCATGTT GTAGAGGCGG CGCTGGCTGT CTAACAATTA TGGAAGCAAG TGTTTGCGGA	9420
AAAAATAAAC CTTCCGTTTC GGTTAGACCT GTACTAGCAA GGTAATAGTC CCCCACTTTT	9480
GTTCCCATTG TTCGCAAAGA GCCAATGAAT CGGTGTGTTT TTTCTTCTAA ATTGTTGGCG	9540
ACTACTAAGC CAGTTCGTAC CGCATTGTTC ACTAACGGAG CATAAAAAGT TTCCGCGACT	9600
GGTTCATTCA TAACTGAAAT GCAATCACCG ATGGCAAAAA CATTCGGCAC TGAAGTTTGT	9660
AAATAAGCAT CCACCGCGAT TGTCTGATCC AGATTCCGTT GAATTTTTTT ATCCAAATAG	9720
GCTAATTGAG GATGTAAGTT TAAAGCAAAA ATCCCACTGT CACAAGAAAT TTCCTGCTCG	9780
CTTGTTTCAA GAACAATACC ATTTGCTGTT TCTTCAATTC CTAGGACCGT TTCTTCAAAA	9840
TGGAAATTA CTGCTTGTTT TTCCAAAGAT TTTTGAACCT CTGCGACCAT TTCTTTATCA	9900
AAATATTTAG GCAATAGATT TTCCAAGCTT TCAAAGACAT GGACCGTTTT CTTCATTTTC	9960
ACAAGAAAT CAATAGCTTC CATTCCAATG GGTCTGCAC CAATGACGGC TACTGTCTGG	10020
CTATTTTCCA GTAATGGAAC AGCCGCTAAA GCACCTGATA AAAATTTATA TTTGAGTAGT	10080
TTTTCTGTTT GACTACCACG AATTTGCGTG GAAACTGGC TCGCGCCTGT CGCTAAAATC	10140
AATTTATCAT AAGAATACCA CTGTTGCTCT TCCTTTCTGT TCCAAGCAAT CAATTGGTTC	10200
TCAACATCCA TTGCCACAAC TTCTCTATTT AAAAGCAACT GAATCTTTTG ACGACGTAAT	10260
TCCTCTTCCG TTATGTAACG TGCTTCGTGT AATTCATTGA TGGTATGATT AAAATACGCA	10320
CTTAAGCCAC CAGATAAATA CCCCACTGTT GCTTGTTTAT CAATTAAAGA AATTTCTGCT	10380
TGGGGATATT TTTTCCTTGA AGCAATTGCT GCGGAAATAC CTGCGAATGA TGCACCAATA	10440
ATCACAATTT TCATTTGTgT CCTCCTCTGT CAGGACCTAc TATACTTCTC TCAGTTTATC	10500
ATGATTAAAA AGAaTGCTAA AgCAACAACt ATAgTTTTCC CTTAAAATAA TTATtTTTGC	10560
TGAGCGCATT GCTTTTCAAG CACTTTCATT TTCATTTATA ATAAATGTTA CTACTGTTTG	10620
AGGAGGAGTT TTCATGTCTG CAACAAAAGA GCAATTGGaA CTATTTTAT TTTATTTATC	10680
TGAAACCCAT ACAAAAAGTT TATCTTTACA TGACTTAGTT ACAAGTCGAC CACGCCCAGA	10740



AGAAGCGCGT	ATTTTATTAA	ATATTAACGA	GGTTTTCAC	TAÇTATCATA	GCGCTCGTGT	10800
TCTCTACACA	AGTGTTCCTG	CTCTTGAAAA	CAACAAGTCT	GAGCCTTTTT	TTCAAGCCTT	10860
CGAAAATTTT	TATTTTGAAT	TAAAACAACA	TTTTTTCAAT	GAAGAAGATG	AAACCAATCA	10920
ATTGAATGAA	CGTCTAGAGG	AAATGAAAAT	TGCCTTTGAA	CAATTGACGG	ATGACTACAA	10980
TGTTCTATAA	TAGAAGAGAA	CAATTCGTTG	AGAAGGGGTC	CTTATGAAGC	AAGAAGAAGT	11040
GATTGAACGT	TTAAAAGAAG	AACTAAATTT	GCCATTTTTT	AATGGTATGC	TAGAAGAAAA	11100
GAACCTATTCT	GAAGCAGATT	ATCAACAAAT	CAAGAAAGAG	TTAATTCAAT	ACTTTGATGA	11160
TTATGTGCGA	AATGTTGAAA	ATTAACAAGC	TAAAACAACC	TTTGATTTTA	ATAAAATCAA	11220
AGGTTGTTTT	ATTATTAAAT	AAAAAAATA	ACAAAAAAC	CACCGGCAGT	TTTGCCGATG	11280
GAAAAGGAGT	ATTTACTCAA	TAAGAGTAAA	ATGAAAAATA	AAAAGGTTGT	TGTTGGTATG	11340
AATTAATAAT	ACAATACACT	TTCTCTTTTG	TCCATAAAAA	AGACATGAAA	AAAATATTAA	11400
ATTCGTTAGT	TATTTTATT	AGTTTTTAT	TTTTTCAAAA	GAATGTTGCT	CATCCGCGTA	11460
ATGGCAAAGG	TTAATTCATC	CGTTGCTTGA	TCCTGCTTTT	CCTCTGAAGC	ATTAAATAAA	11520
TCTGCATCAA	TACCTGTTTC	TGGATCTGTT	GCTAAAATCA	TTTCATTGCA	TCCTGCTACA	11580
TATTCTGTAT	AAGCATGTTG	AAATTTTTTG	TGAATGCCAA	GTAÇTTGTGG	CGTCGGTTTT	11640
AATGTTCCGA	CTTTTTTGAG	CATTAATTCA	TATTTTTTCAG	TACCTTCTTT	AAATTTTGCT	11700
TGAATTTCTG	CAATTCGTTc	GGCTGTTAAA	TCATTCACTT	tATTCTTATC	GATCGCTTGA	11760
CGAACTTCTT	CGTAATATGG	ATyCATGACT	TCGCCAATGy	CTtCTGTkGT	TTTAACGATA	11820
TCATtAATTG	TTkGAACGTA	AAATCTTAAA	TTTGGTCTCA	TAACCATCCC	TCTTTTCCCT	11880
TTTCTCTtTC	ACTCCATTTT	ACACTTTTTC	TATTCCATCT	GACAAGTGGC	ACTTTTTCAG	11940
TCTTCATTCT	TTTCATTCTT	AACAGAGAAA	ACATAGTGAG	GCATCGTTTT	GCCACGGTAT	12000
GTTTAAATGT	ATCGATCCCT	AACAAGCATC	TGGTTTCGAA	TAGCGACATT	CATAGAGGCT	12060
AAGTTCGTAT	CACGAATCGT	TGAAATAACT	TCTGGAAAAT	TTAATACATC	GAATGCGTAC	12120
TGTTTTGTCTG	CTGCGGCAGC	TTCAATCGCA	TACCCTTTGT	GCCATTCTTT	CACACATAAA	12180
TGATAACCAA	TTTCTGGATA	GCTCTTACCT	TTTACTAATT	GATTGGTGAG	ACCACACTCG	12240
CCGATGACTT	CGCCTGTTCTG	CTTCTTCTCT	AACGCCATA	AGCCGTAGCC	GTATTCTTGA	12300
TAGGACTGTA	AGTTCCACTT	TAGCCAGTTT	GCGACCTCTT	CATCAGAAAA	AGCGTGTTCA	12360
TAGGCATACA	TCACATTTTC	ATCTTGCAAA	AATTTTTTCA	ATTCCACTTC	ATCTTCCGAC	12420
TGCCATTCTC	TGACAATAAG	TCGTTGTGTT	TCGAAAATAA	TTTCCGTTT	CAAGCCCTTT	12480
CACTCCTTAC	ATAAAGAAAA	ATTGACTGGT	TTAAGTACCA	GCCAATTTTC	TTTTCTTATT	12540
TATTAAGCAC	CTTTACCTGC	TTTACGTTGG	AAGAACTTCA	CAATTTCCAC	AATTGGGATG	12600
ATAGCAAAAG	AAGTACCTGC	TACAATTGCC	CATTGATATG	CATCTAAATG	TGTGACGCTG	12660
AATAAATCAT	TAAATCCTGG	AATAACAATC	GTTGCAGCTA	ATAGCAAGAA	CGAAACAAGA	12720

ATCCCATAGT TAAATGATTT GTTGCGGAAT AAGCCAACTT TGAAAATTGA TTGATAAATC	12780
GACTTCACAT TAAAGGCATG GAACAATTGA ATTAATCCCA AAGTAGCGAA GGCCATGGTT	12840
AAAGCATCGC CGTGTGTGTA ATCATACAAC GcTTGCGCCG ATAAGTTTGT GTTAGCTGCG	12900
GTATGTGCTG GGAACATAAT TGACATTTTG TACACAATCA AAGTCAATGC CCCTTGCGTA	12960
ATCCCTTGAT AAACAACACT GCTTAAACG CCGCCAGAGA AGAAGTTCGA TTTTTTACCA	13020
CGAGGTTTAT GGCTCATCAC ATCAGGTTCT GCTGGTTCCA CACCTAACGC AATCGCTGGG	13080
AAAGTATCCG TTACTAAGTT AATCCACAAT AAGTGAACAG GTAACAAGGT ATCCCAATTG	13140
AGCATTGTTG CAATAAACAA GGTTAATACT TCCCCTAAGT TAGCTGAAAG TAAATATTGA	13200
ATCGTTTTTTT GGATATTTGA GAAGACTTTA CGTCCTTCTT CTAAGTCCAC AATAATTGTT	13260
GAAAAGTTAT CATCGGCAAG GACCATATCT GAAGCACCTT TTGATACTTC TGTTCCGGTA	13320
ATTCCCATAC CAATCCCGAT GTCAGCAGCT TTTAAGGCTG GGGCATCATT TACGCCATCA	13380
CCAGTCATCG CAACCACTTT TCCTTCTTGT TGCCACGCTT TAACAATCCG AACCTTGTGT	13440
TCTGGTGAAA CACGAGCATA GACAGAGTAG TGTCCACGA CTTGTGCGAA TTTTTCATCT	13500
GACAATTCAT TTAATTCAGC GCCTGTAATG ACTGCATCGT CGTCGCCTTC TTTGATGATG	13560
CCTAAACGTG CCGCAATTGC TTCTGCTGTA TCACGGTGGT CACCCGTAAT CATAATTGGA	13620
CGAATACCCG CTTCTTTCGC GACTTTTACA GCATCTGCAG CTTCTTTTCT TTCTGGGTCA	13680
ATCATACCGA CTAAACCAGC AAATGTTAAA TCTTTTTCCA CAAGTTCAGA ACTCATTTCTG	13740
GCTGGAATCG TCTCAACATA TTTATAAGCC ATTCTTAAA CACGTAAGGC TTGTTTCGCT	13800
AATGAGGTGT TGGTTTTCAA AATTTCTTGA CGTTTTGTTT CATCTAATGG ACTTGTTTCT	13860
CCATTTGATA AGATTTCCGT ACAACGTTT AATAATTCAT CTGGTGCACC TTTGACAGAA	13920
ACTAGGAAGC CACCTGTTTT TAATTCATGA ACAGTAGTCA TTAATTTACG ATCAGAGTCA	13980
AAAGGAATCT CTGCCACACG AGGTTACAGT GCTACTTTTT CGGTTACGTT AAACGCATGA	14040
TCTAAACCAA ATTGTACTAA GGCAGTTTCT GTTGGGTCAC CAATTAATGA ACCATCTTGA	14100
GCGATTTTTG TATCATTCGT AAAGTTCATG ATTTTTAAAG CCATATTATC GGCAGGGATT	14160
TCTGTGGAAG CAGATAAAAC TTGACCATCT GTATACAACG CTTCAACGGT CATTTGATTT	14220
AAGGTAAACG TACCTGTTTT ATCTGAACAG ATAATGTCTG TACTTCCTAA TGTTTCAACG	14280
GCTGGT	14286

## (2) INFORMATION FOR SEQ ID NO: 163:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6723 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

CAATGCTCAT TTCCACTCnT CCTAAATCTG TTATTGCCAG TAGTTTTGCT CTAACTTTTT

TAAATaATGk GaATAGTatC CATGCTGTCG aACCTCAAAT TCTTTTAAGT AATCtAATGG	120
AATACTCyTA CGtTCTAAGT TTTTTTGCAT ATTTAAATAA TCATCAACAT CCaCTAAAAA	180
TACTTGATGA CTTTTGTGA ATTGAATTAA TAAAAATGCT GTTCCACCGC TTAATTTAA	240
AnCCTTTAA AAATTTTTTT GGTGATCTTT TAACATTGGT TTATTTCCAA CATAAAATGG	300
AAATGCAGTT TTATTTTCAG TAGATTTGCA ATCGAAGGCG ATTGGCCGTC CCTTCAAATG	360
TCCGATAAAA TCACATCCTG TCTTGTTAGT TGGAACTACT ACGGGTTTTT CACCAACCTT	420
TATAGTTTTA GTTCCATTAG GTATTTTGC TACTGTGCCT TTTCTGTTAC GACAATACCA	480
CTCGTTGGTT tGTTCAATCA TTTTTTCAA CTGACTCCAC GCTTTCATCA CAGCACCTCT	540
CTCTTAATTT CCTCTGGAT TTAACCAGTT ATGCATTGTA ACTTTTGGTA TTGAATCGAC	600
AGAAATCTCC TCTTGATCAT TTCGCATATT GTATTCATA ATTTTTTTTA TAGCTTGATT	660
TTTTGTCTTT ATTCCCTTAC GTTCCAGCT TAACAATATG CGGTCCATAT ACTTTAGACT	720
GAATGCCTGA TTAAGAACTG CCTCTTAAAG GGCTAATTGA ATAAGGTCAT CTGGATAACC	780
ATCTTCTGTT TGCCACTGTT TAATCATTTT CATTTGCaTA GGCGAAAGTG GTctCCCaaa	840
ATTTTGTTCg ACGGcAGCGT ATATATTTAA tTTaTTTTTA TCTGTTTCTG GTGAAAAAGA	900
ATTATTTGTG TTTATGTTTT GTTTATGTTT AATTAATGTG CCACTGTTCTG TTGACTGCGT	960
TGTACTCTTT AGTGTCACTT TATCGTTACT CTGTTGTGTA CTCTCTTGCG CACTATTTAA	1020
CGTATAAAGT ACGCATACCT TGTAAGAAGT AGCTTTTCGA CCATTACTTT TAAAATCAAT	1080
CAAACCTAGT TGCTTTAACG CGnTTCTATT TTTATTAATT CCTGAGCGTG ACAAACCTGA	1140
CAAAGATTCT AATGTTGCAT TAGCTGCTGT AAACCATGTA GCCCATCCTG CTTTGTGTGT	1200
TATGGACATT AATGCACGCC ATAAAGCAAT CTGACCTGAT GAAAGCTTTT GTTTATAAAG	1260
CAAATAATCG TCAAACGCAA GAATCTGTTG TAAATAATTC AATGTCACAC CTCCACTCTC	1320
TATGCAATAA TTAtTTtTT ATTGGTTATC TtTTCTATtT CTtCTTTAAA TTTTCTGGa	1380
TCACTATTCT TGTCGCTTAA ATGGATGAGA TATATTTCTT CTGTTTTTGT TAAATCAGTT	1440
GACTGAAAAA ACTTTTTGCA GGCATCTATG CTCATATGCG TTCTTAAAT TCGATCTTGA	1500
ACACTTTTTG GCAGTTTGCT CTGTCTTACT AACTTGATAT CATGATTACA CTCAACGAGC	1560
CAATGAGTGA CGTCTTTAAA CGTTTTAGGT AAATAGTTAG TATCTGTAGC AAATACTATT	1620
TTCTTTCCAC TAGGTGAAAG AATAAGAAAG CCCAGTGGTT CTCTCGCCCG TGCCTTTTTG	1680
TCATCGTGAA TAGTGGCAAA AGGTTTAACT AACCAATCAC CAATTTTCTG TTGCTGGTTA	1740
GCTTTTAAAA TATGTGACCG TCTATTTATA CCTAATCCTT CTAAAGTTCC TCGTGAAGCC	1800
CACACATCAA ATCTTCCAGC GAGTAAATA TCATTGATAT ACTTTGAGTG GTCACCATGT	1860
TCATGTGTTA CAAGTAAACC TTGAATATTT GAAAAATTAA TACCTTGTTT CATAATATCT	1920
TTAGGCTTTA ATCCTGCTTC TAACATTAAG GAAGAGTTTC CATCTGCAAG TAAATAATTA	1980
TTACCTGCAG ATGAAGAACC TTGTATATTA ATTTCAATCA TTAAGGCCA CGTCCATCTG	2040

TTTCGCTAGC AATATTTGTG GTTAATTCAT CAAAAACGC TGTCTGGGT GGTTCGTTTT	2100
CTATTTCTTG ATTCGTTGTT GCTTTTTGAA TCGGCTCTTC TGGTTCGATG ATAGTCTCAT	2160
CTGCTGCAGT TTCTTTATCG AGTTTTACTG CTTCTTGTTT GACTTCTGTT GGTTCCTCAT	2220
CAAAATCGAA CACTTCTTGT GCTGTTTGCT CAGTGACATC TTTTCTAACA GATTAAACCA	2280
CTTCATCATC TTGCGACTGA TACTGCATTG CGACATAGGC GTTTTCAAAA TTTTAGGAA	2340
TTTTCTTAAC AATGktGTTT CTCATTTTAC GAACAATCAT GGATTCTCTA CTTGTGGAG	2400
ATTTCCATGC GGGACTAATA TAAGTCTGGT ACTCTTCGCT ATCCAAGATG TCATCTAAAG	2460
ACATTGTTTT CAAATTGTTC ATGATTTCTT GCTTTTTATT TTCGATTCTT GTTCTCTGAC	2520
TCTGAGTTGC TTTGAATTTA TTTTCAGCTA GTCCAAACGT CTCATTCATC AAATTTTGGT	2580
TGATGTGGGC AATTAAGTTT TTAACAACGT CTGCACGCTC TGAAATATGG AACTCAACAG	2640
ATCCGTCTGT CATTTCAACT GGATAAACAA CTCTAACAC TTTTCCTTTT CCAGTGGGTC	2700
CCCACTCTGG ATCAGTCACA GACAGCCCTT TGTATCCAGG GTACGAAAAA TGGTCCTCTT	2760
CCCGAACTTC CCAATGTCGA TGAACATGCG CCACATTTTCG TCCGAACTTT GAAAGAATGG	2820
CGTCATTACC ATCTCCTTCA ATTCCCATTT CAATCTGTTT AATCCAAACA TCTTTCGGCT	2880
TGCCCCGTGTT TGGATCATT CTTTTCTTT TAACATTTTCG GGTTTGAAAA TATaCTTCGC	2940
GAGGAATCGC CGATGCATT CTTGTAATG CTGCGATGGT CATTAGCGTT TCTGTTATAT	3000
TCGTTGAATC TACATCATTA ATTGACAAAC TGGCATTGTG aAGCATTGTA TTAAtTGCTT	3060
GAAWTGAGCT GATTACGCAT TGTTTTTGAT ATTCATTATC ATTAATTCCG TTGCTTAcAA	3120
GCTGCGAttC AACTTGTTGGT AAAAACGTAT CGTTGATTTT AGTTAATCTA TTCTCGAAAA	3180
TTTTATTTTT TGCAATTTCA TTGGTCATCA ATATCTATT CTTTCTCTGT CAAAGCAGCC	3240
ATCCATACTT TTTTAATTTT GTCGGGACAA TGACTCATAG CATCTTTCCA AGTTGGCCAT	3300
CTTCCGTTTT TGTCATAGAA TTTATATTGA TATGTTAAAC TTTGCTGATT ATGTGGTTGC	3360
TCTTCATCAT GctTAACTGC ACAAATTTTCG CAGGTTCCCT CCGGAACGTT TCCGATCATC	3420
ATAAACCCAG TTTCTGCTTT TAGAAATCCC ATCTATTTTG CACCTGCCTT TTCAATTGaA	3480
TGCTGTTCAA TTCTAATTTT TTTAyCTTTT TCGCTTACAT ACATCGCAAT GACTTGTGTA	3540
TCTACTTGAA CAGAGTCTCT ATTGTGATTA GTCAAACCTT CAGCGTTATC GATAAAAATT	3600
GGAACGATAT AGCCTTCCTG TTTCATTAGA GTGTTGGAGA CATCTAATCC AGCTTGATA	3660
CGGCTACCGT TATTTAGAGA ACTGAACGGT ACTCCGTCAA TCATCGGTTT ACATACTGCT	3720
TCATTGAGTC CTCCATCTTC AAAGAAATCG AACAAATTCC ATTTAACAAAC TGAGAAATGA	3780
CTATTTATAA TTTCTTGACG CATGTTCTGT TTTGTTATAA AGAATTCTTC AAATAGTACT	3840
AATTTTTGTA ATACTTCACC TTTCTTATGA GATAATTGGC GCTCTTGTTT ATTGAAATCT	3900
TCAATAATCG ATAGCTGTCT TTCATACTCT TTTAGCAAGG CTAATTTTTC ATCAATCATT	3960
GCGATTCTT TATCTATTTT TGTAATCTCA CTTGTTTTAG CGGAAGTTTG TTCTAGAATA	4020

GCTTCGTTAC TTTGAGTAAT GTACTCTTGC AATTTTTTTT TTTCTTTAGT GATAGTTGAA	4080
TATTTTTTCAG TTGCTTCAAA TGGTATTTTG TCTAATTTTA ATGAGGAAAT CTGCTGCTCT	4140
ACATCAGCTA AGTTTTCCCT CACATCTTCC AAGTGTTTTT TAGCAATGTT ATATGCTTCT	4200
GTTTTAATCA ATAGTTGTTC TTTAACTCG CCGATCCCTT TTTTAGGgC TTCGCGATCT	4260
TTGTTATTTT GAATGCCTTT TTCTCGaATT TCAGAGAGCT TGATTGCTT ATCCGCTTCA	4320
AACTGGGCTT TTATTTCTTT ATTTGTTAAC TCAATTTCTT CTGCACGTTT TTGCTCTTCT	4380
TCCTCATGAT GTCGTTTCAT TTCATCTTGa TCCTTGACAT CATATGGACG ATTACAGTGC	4440
TGACAAACTA AAAGATTTTC ATTGAAAGAT AATTTGGTAT AAAC TAGACC GCCTGTAAAT	4500
TCTTCAGCTT CTACTTCATC ATATTTGTCTG TACAATTCTT CATGCTTTTT ATTTAATGCA	4560
ATCAGTTCAT TATCTTTTAT AGAAaCCAAA CGTTCTGTTA CATTAAAGACT TGATTCTTCA	4620
TCTGCATACG TCTTCTGGGC TTTATTGAGA TCAGCAAAGA GTTTGTACTT ACCTTGCTCA	4680
ATGCCATTAA TGCGTGCATT CTGTGCATTA TCATGCTTCA ATTTAGCAGC AGTTAATTCT	4740
TCTTGTTTTG TATTAAGACT TGCAATTAAT TCTGAAATAT TACCACCATT TCTAATAGTA	4800
ACGAGCTGGT TTTGATGTC ATTTTCTTT AATGTCAACT CGTTGCGAGT GGTTAATAGT	4860
TGCTCTTTGT TGATATTTTC AATATCTGGc AATGCTGCTT GAATACCTTC AATTTTTACA	4920
GGAATATTTT TCAGCGTTTC GTTGATcGCT TTTGTCTTG AAGCACACGC TCACGAGCTG	4980
TTTTAATGTC ATCATTACCA ATAATTtCTT TTAATTGGTG AATTGATGGC GATTCGTTGA	5040
TAATCTCTkC ATCTGTCTTG CTGCCAAAGT AtTCAAAAAG CTTTTGACGT CGTtCATCTG	5100
CCACTAACTG CTCACAAAAA TAAGTCACAC TAGTCAGATT CTTAAATGTG TCTTGATCTA	5160
ATACTTTTTT TACTTCATCG TCAAATGCTT TCTTAGTTGT GGTCTCTAAT CCGTCTACGA	5220
GGTATTTTGT AAACATCTCA TAAGACTTGT GTTCCGAATT TCGTTTGATA ACCTCTTTGT	5280
CTCCTCGGAC CTTTTCAAAC TCCTTTGCCT GACCATTGAT AGTCAAAACG ACAGTTACTG	5340
ATGTCTGTTT ACCGCGAATT GGCTCGCTAT TTTATCAAG TGGACGCCAC TGAATCTTAG	5400
TACGTCTTTT TGAATCTTTG TTGAACAAGC ACCAGAGAAA aGCATCaTAG ATAGTTGTTT	5460
kGCCTGCATC GTTGTCTCCA AAAATATCAA TACTTTTACC ATTTGGTTCA ATCATCAGGT	5520
CCGATATACC TTTAAATTA TGAATACGGA TTGACTCCAA GCTAATGTTT TTCATACAAT	5580
TAATCTCCTT CAATTTTCGAT ATTGTCTTCT AAGATTGCTC GTATAACTGT TTCTTTGCGT	5640
TTGTTACAAA TTTTCAGTAGC GACATCATAC CAAGCACTAC TATGATTTTG ATTCAGCTCT	5700
GTTGTAAAT CATATCTATA ACTTTTACTA ATTATTGGAT CAACTAACCA ATCCATACGT	5760
GCAATTAAGG CTGTAGCGTA TAATTCTCGT TCTGCGCAAT TTTTCAGTCGC GTTTTTAAAC	5820
CATTCTTCAA ACCGTTTTGA CATGGTTAAT TTATTTGACA CATTATCCAT TTCATTTCTT	5880
CCAGATTTTG TGTTATAATT CTCATGTATA ATTTTTGTAT GGGACTTAAT CGTTTGTGGA	5940
CGAGTGAGTC TCTTTTTTTG TGTAAGCTAA ATATTTTGCA TCGTCGTATT TCATGAACCA	6000

918

AACAACTGCA ATTGGTCCAA TAATCAAAAG TAGATAGcTC GCTGGAACAC TATTTTAAAT 6060  
 TaCTAACCCCT AATATAAGA CAAGAATAAA TGATCCTAAA ATTCTAGCTt CGTATAAGCT 6120  
 TTnTGGTTTT CTTECCTCA TTTATTTTCC ACTCCTATTT ACCCATAGCA CTCTGATAAA 6180  
 CATTAAAAA AGTGCTATA TAATCGCTGT TAAAAATTGT GATTTAATCA AAGATAAAAT 6240  
 AAACACAAAG AATAAATTTC CTACGGAAAG TGTCAATCATT GTTTTAATTG CAAGTTCCTT 6300  
 ATTCTTCACT AGAGACCCCTC CTATATAAAG TTTCTTTTCAT ACCATTCTTC TAAATCATCG 6360  
 ACTTCTATTC TTACCAATCC ACCATTTCTT TTAGCAGGTA ATGGATCAAT TTCTCGTTTC 6420  
 ATCCATCTTG TAACAGTAGC AGCTGAAACA TTTTTTCTTT TTGgCTACTT CTATTGCTTT 6480  
 CAAGGTGGn AATTTCCTTG aATTAACTTT TCGaTCKTTT TTgaTTTAAA ATCAATAACT 6540  
 TCTAGTGcCA TTKGwATyCT CCTTCCTCAT GtATCCTAGT kGTTCCCAAT ATGGAAAACG 6600  
 TTGCTCACTT AAATAACGAA TATCGATGAA AGCAAGATCA CACAAGCTAC TAAGTAGTGT 6660  
 GATTCAACAA TTACTIONCA AATAnCCATT GCGnAGCACT ATTTAGAAAT ACTGTCAGAA 6720  
 AAA 6723

## (2) INFORMATION FOR SEQ ID NO: 164:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27360 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

TAAGCGAATT GGACATCATA TGTCAATAAC AAATGTTGAT AAAAAAGGTC AATTAAACGC 60  
 TATTTTGAAA AATAAAATCA GCTTTTAGAT AATTTTGCCA AAAATTTTCGC TTAATGAGTA 120  
 TTAAAAATAT TTTTATTATT TATGTGCTCA AATATTTTTC AAAAAGAAAA AATTGTATTA 180  
 AGGTAATAGT AGAAGGGAG GAGATTTAAT GACAGAACCA AACTTAAAAG ACCCAAGAAA 240  
 GTTATTTTAT AGTGATGGAT TTCCGCAACA AGATCAAGAA ACTCCCGCAT TACAGGACAA 300  
 AATGGTCCCT AAACCAGATT GTGGGGAAGA CAGCTATGTC GGAAATCACA AATTAGAAAA 360  
 TCGTCGGGTA TTAATTACAG GCGGTGATTC TGGGATTGGA CGAGCTGCGG CGATTGCTTT 420  
 TGCTCGTGAA GGCAGAGATA TCGCCTTACA TTTTTCCTCA GGTGAAGAAA AAGACGCCGA 480  
 AGAAGTGGCG CACTATATAC GAGAAGCTGG CCGAAAAGTC GTTCTTTTGC CAGCAGATTT 540  
 AAGAGATAAA CAGGCACCTG AAGAATTGGT TGCACAAGCT CATGAAGCAT TAGGCGGCTT 600  
 GGATACCCCTC GTTTTAAATG CAGCCCAACA AATTTCTGT GCCGCTATTG AAGAATTACC 660  
 AATGGAACAA GTGATAGACA CTTTCCATGT CAATATTATT GCGATGTTTG GCATTGTCAA 720  
 AGCAGCAGTG CCTCACTTGC CTGCCGGAAG TAGTATTGTG ACAACGACTT CCGTCCAAGC 780  
 CTTAATCCC AGTGAGCATT TGTTGGATTA TGCAGCAACG AAGGCATCAA TTGCCAACTT 840

TACCGTTGGT TTAGCAAAAC AACTGGCACC AAAAGGCATT CGGGTAAATG GCGTGGCTCC	900
AGGACCAATC TGGACACCGT TACAGTTAGA TCATGGACAA CCAATTGAAG AGTTGCCAGA	960
ATTTGGGCAA CATTCTTTAC TGGAACGTGC GGGTCAGCCT GCTGAATTAG CCCCTGTTTA	1020
TGTTTTCTTG GCGTCCAATG ATGCAAGTTA TGTTACGGCT CAAGTTTATG GCGTAACTGG	1080
TGGAGAAGCA ATTAATCTTT AGGAGGGATG TTCCGTGAAT AAAGTTGTTA AAATACTTTT	1140
GCTTCTAGGT TCACTGATTT TAATGGCATT ATTTGGCGCG GTGGCCATCA CCACTTGGTC	1200
AGCCGATACA GCTTGGCAAC CACCAGATAA TGTCCAATGG TTCTTGCAAG ACAACGAGTA	1260
CGCACGTCAA GCAATCTTTT GGGTAGCGCT GGTTTTACTA GGAATTATGC TCTTTCTCTT	1320
TTTATTCACT CTTTTATTCC CAAGAAGAAA AGGCACGTTT GAATTAAAAG AAGAGCAAGG	1380
AAAGTTAGTG TTACACCGTA AAGCCATCGA AGGATTTGTA CGCTCTAGCT TGCAAGACAG	1440
TGACTTTCTG GAAGCACCAA CCATTCGTGT AAAGGCGACA AAACGGCGCA TTAAGGTCAA	1500
TGTTAAAGGC ACACTGAAAC GAACGGCAGA TTTAATTGGG CAAACAGAAG CTTGGTCAAA	1560
AACAATGGAA GAACGGCTTG CTCGTTTGGT GGGAACTGGT CACAAGATTA CGATTGATGT	1620
TCATTTTGAA GATATTCAAT CAATGAAACG ACCACAACAA CAGCAAACAC AAGCACGTGT	1680
GGAGTAGGAG GGATGTACAA TGGTTGAATG GATAAAACAA TATGCCGTTT CGTTAATTGG	1740
TGGCCTAGTC GGCTTGGTTT TTGCAATCTT GTTTCGTGCA ATCGGCTTCT TTAAAACATT	1800
GTTAGTGGTA CTCCTGACAG GACTAGGTAT TTGGCTAGGT TTGTATTGG AACAGACAGG	1860
CATTGTCAAA AATTATTTTC ATCATAAATA GGAGGTTTTG GCTATGGAAA ATGCAAATGT	1920
ATCGATTGTT AAAGGTGAGT TGACGTTTGA GGATAAAGTA ATCCAAAAAA TTATTGGCAT	1980
TGCTTTGGAA GAAGTCGACG GTTTGTTAAC TGTGGaCGGT GGTTTTTCT CGAATTTACG	2040
AGATAAAATG ATTAACAGTG AAGATGTTAC AACAGGCATT CACACCGAGG TCGGCAAAAA	2100
ACAAGTCGCA GTGGATATGG ACATTGTAGC CGAATACGGG AAAGATATCG AGACCATTTA	2160
TGATCAAATG AAAGAGGTTA TTAATCGCGA AGTCCAACAA ATGACTCACT TGGAAGTAAT	2220
CGAAGTCAAT GTCAACGTAG TAGATATCAA AACACAAGCA GAGTACCAAG AAGAAAGTGA	2280
AACGrTCAA GATAAACTTG GCAACGCCGC TGAAGCGACA GGTAGTTTTA TGTCAAAACA	2340
AACCAACAAA GCCAAACATG CGGTAACATA AGGCAAGACA AAAGTCAAAG AAAACACGGC	2400
ACCTCGTGTT CAATAATTGG TTATTATTTT ATTAATAATA AAGCAATAAT AGAAGAAACT	2460
AGGAGGAATG AACCATGAGT AACGAAAAAT TCAATAATGT ACCAAAACCA GCAGGAAACG	2520
GACCACACAC ACCAGAAGAT CATGCAATCA AAGGTGAATT AACATTTGAA GATAAAGTCG	2580
TTCAAAAAAT TATCGGTTTA GCCTTAGAAA ATGTCGATGG TCTATTACA GTCGATGGCG	2640
GCTTTTTCTC AAATATTGCT GAAAAATTAG TAAACACAGA TAACGTCACA GCGGGTATTG	2700
ATACCGAAGT TGGTAAAAAA CAAGTCGCTG TTGATATGGA TATCGTGGTA GAATACGGCA	2760
AAGATATTCA AGACATTTAC GAAAAAATGA AAGAATTAAT CAGTCGTGAA GTGAAAAAAA	2820

TGACCCACTT	AGATGTGATT	GAAGTGAATG	TTAACGTAGT	TGATATCAAA	TCAAAAGAAG	2880
AATACGAAGA	AGATAGCGAA	ACGGTTC AAG	ATAAAGTCAC	AGGCGCTGCC	AAAAGTACCG	2940
GTGAATTTGC	TTCTGAACAA	ACAGAAAAAG	CAAAAAAAGC	TGTGAATAAA	GGAACAGAAC	3000
AAGTCAAAGA	AAATATGGAA	CCACGTGTAG	AATAAACAGT	TAAAGGGAGG	AAACAATCAT	3060
GGGCTTTATT	TGGGCATTAA	TTGTCGGCGG	GGTCATTGGG	GCAATCGCTG	GAGCAATTAC	3120
TAAAAAAGGA	TCATCGATGG	GCATTATTGC	CAATATCATT	GCAGGGTTAG	TTGGTTCAAC	3180
AATTGGTCAA	GCCATTTTAG	GCACATGGGG	ACCAAGCTTA	GCTGGGATGG	CTATTGTGCC	3240
ATCGATTATC	GGGGCAGTGA	TTTTAGTTGC	CATCGTTTCA	TGGATTTTAG	GACGAAAATA	3300
GAGAAAGTTT	AAAAAAATAG	GGCATAAGTC	GTTTCGGCTT	ATGTCCTATT	TTCTTTTTTC	3360
TATGTATGAG	GAAGTTTCGT	ATGAAAAGTA	AATTTTGAGC	AATTAGCCAA	GAGCTTATTT	3420
TAATTAAAAT	CATTAGTCAT	TTTGGCTTTT	CAACAAAAT	CAACAACATT	TGTTGAAGTT	3480
ATTTGTAGAT	AACGTTTATC	ATTCACAAAT	TTACATTTT	TTGTAACGGA	TATCTCGTAT	3540
AGTAATAACT	GCAAATGAAT	TTCAGAAAAT	TTTTTGATAA	GTCATCAAAG	TGAGGGGATA	3600
GCGGTCTTTT	ATTGAAAAAT	AGCAACACAA	AATAACAGA	ACAAAAGAGA	GAGGAATTCA	3660
AAGAATTATG	ACAGAACAAA	CACCAACAAA	GCTTTTGAAT	AAAGGCTTTA	TCAGTATTAC	3720
AGTGATTAAT	TTCATTGTCT	ATCTTGTTTA	TTATTTATTA	ATGGTTATTA	TTGCGGTGAT	3780
TGCGCAAGAT	AGCTTACACG	CTACTTTAGG	TCAAGCGGGT	TTGGCCTCAG	GAATTTACAT	3840
CATCGGGACA	TTGCTTGCAC	GACTTTTCAT	GGGGAACTC	CTTGAACGTA	TTGGGCGCAA	3900
ACAAGCTTTA	CGTTATGGGG	CTTTGTTTTA	TCTTTTAAAG	ACTGTTGCGT	ATATGTATAT	3960
GCCAAGTATG	GGCATTCTTT	ATCTGGTCCG	TTTTTTAAAT	GGGTTTGGTT	ACGGAACCGT	4020
TTCGACAGCG	ACCAACGCAA	TTGTTACCGC	CTATATCCCT	AAAAACAAA	AAGGGGAAGG	4080
GATCAATTAT	TATGGTTTGA	GTACTAGTTT	AGCAGCCGGC	ATTGGGCCAT	TTATTGGGAT	4140
GTTATTGTTA	AATGTTTCCA	ACTTCCATGT	CATTATTAAC	TTTTCAATTA	TTTAATTTT	4200
GTTAACAACC	ATTGCTTGCT	TCATTTTCCC	AGTGAAAAAT	ATTGAGTTAA	CACCTGAACA	4260
TCGTGAAGCT	TTATCTAAAT	GGAATTTTGA	TAGTTTGTG	GAAAAGAAAG	TTTATTTAT	4320
TACATTTATC	GCCTTTTTAA	TGGGTCTTGC	TTATTCAAGT	GTCTTATCTT	TCTTATCTTC	4380
TTATGTCAAA	GTAATCGATT	TAGTTGACGT	CAGCACCTTC	TTCTTTATCG	TCTATGCGGT	4440
CGTGATTACC	TTAACACGTC	CATCGACAGG	TCGGATTTTC	GATGTCAAAG	GTGAACGCTA	4500
CGTCATGTAT	CCGAGCTACA	TTTTCTTAAC	ACTAGGCTTA	TTCTTATTAA	GTATGACTAC	4560
GTCTGGTTGG	ATGTTATTAG	TTTCAGGTGG	GTTAATTGGC	TTAGGTTATG	GTACTTTCAT	4620
GTCAATGGC	CAAGCGGTTT	GTCTTCAAGA	AAGCCCAAGT	CCTCACCGTA	TCGGGATTGC	4680
TTTATCGACT	TATTTTCATCG	GTTTAGACCT	TGGATTAGGG	GTTGGTCCGT	ACGTTTTAGG	4740
TGAATTACGT	AACTTTATGT	CATTCCAACA	AATGTACTTC	TTAGCAGGCT	GTATCCCAAT	4800



CGTTTGTACC	ATTTTATATA	TGGTCTTCCA	GAAAGCTAAA	AATGATGCCA	AAGATTTGTC	4860
TCTTGAAACG	ATTGAAGAGA	TTGAACACGG	CTCAGAACTA	TAAATGGACA	AGGAGTGATT	4920
TCAATGCCTA	ACGATGAACG	GATTTTAAAC	GTTTTACAAC	AAACACTAAC	GGAAATTCAA	4980
CGCATGAACA	CGCGTTTAGC	TAAAATTGAA	ACAGAGTTAG	CCGATGTCGA	TCAACATCAC	5040
GCGACTTTGG	ATCATTACTT	GCAAGAGCTA	GAATTGGTGA	CAGAAAATAA	CAACGTCATG	5100
TAAAGTGAGT	AGTTAGTAAA	TGAATGTTCT	ACGTTTCGGT	ATCACTAGAG	TGGAACATTG	5160
AAAATTTTCG	TAAAAACGT	CAAAAGGTAG	CAGTTTTTCT	GCTACCTTTT	GGCGTTTTTT	5220
GTAATTTTTA	TCACTTAAAA	TGTTTCACGT	GAAACTTTT	AAAAGTGGCG	TGGACTTTGT	5280
TTGGTGACAA	TTTGTGTATT	TAAATTTTCT	AAGCTTTGCG	CTAGTTCACC	AATGCTAGCA	5340
GAAAAATCAA	CATAATAAGT	GGGAACCGTT	GATTTGACAC	TAACCTCTGA	AAAGATATCT	5400
GTAATAATTA	CATCGAAAGG	CTTTTCTGGT	GAAGCGACAA	AAATTAATTT	TTTATCCACG	5460
GCTTTACTTA	GCTTTTCATA	AAGGGTATCT	AAAATTAAAT	TACTTTGATT	GGTCCTTACT	5520
AAGACACGAA	TCGGTGGTAA	AGATTTTCGT	TTCAATGTAC	AAAAAAGAAT	AAACAAGCAG	5580
TCAACTAACA	ATTTTTCATT	GCCAATCAAT	GTCGGAAGCA	TTTCAATTGT	AAGGCGCTTC	5640
ACATCTTTGT	AGAGACTTGG	ATTTTTTGAA	GTGCTTCTTT	TAAAAAGCTT	TCTTGCACAA	5700
ATGGATTGGC	AATCGTGACA	AAGGTCTTAA	AAAAGACACT	TAAATTTTGG	ATAGAAGTAT	5760
ATAGATAAAG	ATACTCGTGA	AATGTCAATT	TAATTTGATA	AAAAGACGTT	AGTTGATAAA	5820
TGATCTCTTT	AATGGGTGAG	CCTAATAGCT	GGACCATATG	CATCCGTGCT	GCAGGAATTT	5880
CCAAGACATC	TTGTGTGGCG	TA <sub>g</sub> cTGaCAC	TAATCGTCAA	TAAGAAAAAG	GCAAAAAGGC	5940
CTTCTTTTTG	TATAAATTTT	GTTGGAACCT	GGTGTAAGGC	TTCTTCTTTA	AAAAACAATT	6000
CAAATTCACT	CAGATTAACA	AAGGGATTAA	CAATCGCTTC	TTCTCCACG	AAAGTTTCTT	6060
TGATAAAAAA	GCCCTCTTTC	CAACGCGCTA	AAATCACCGC	AGAAAAAATC	AAGTCTTTTG	6120
TAAATCTTG	TGTTGTATTG	AAAGAAACAG	AGTGTTTTAA	ACGTCGAAAG	AGACTATGGA	6180
CTTCTTTTTT	AGTAATCGTG	TGCAACGGCC	AATCCCTTTC	AGAGTAATTG	TTCCAAAAAA	6240
GCATCATAAA	GAAATCCGG	ATATCTTTTT	CTTTGCCAC	CAACACATTG	GTACGTTTTA	6300
AATTGATTC	AATCTGATAT	TTTTGGAGCC	ATTGGTTTAA	TTGTCCAATT	TTTCGTGTAA	6360
AGGTTGATGA	AGAAATCTCT	AGGACTTTTT	CAAAATGTTT	GCTATCAATT	GCATTATGGT	6420
TAAAGGCATA	GGTCAGTACT	TGATACGCGA	AACTTTTTTT	GACAAAGTAG	GGCATCAAAA	6480
ACGACAGTTG	GAAATAGGGT	GCAAAACCAA	ATGTGATTTT	TTTGGCATT	GCTGATAAAG	6540
AAAATAAGTT	CTCCAGCCG	TTTTTTTCAA	TTTCTGTTTG	GAGCAAAGCT	ACAGATTGTA	6600
GAATGGTATT	TTGTTGCATT	TTTAATTGAG	TCGCTAAAGT	AGAAATCGCC	AATGTTTGT	6660
CTTCTGCTTC	ACGACGTTTC	AGTAACTCTT	CTAAAAGAGC	AAACTTTT	CGATAAGGTA	6720
CATCAAGTAA	CCCTGTTAAC	ATCCGATTCC	CTCCAAACTG	TTTTCCCCAC	GCAATTTAGC	6780

AAAATAAATT TTGCGAACTA CTTGTTTTAG GTTATAGTAT TCGTCTAAAT AAGGAACGGG	6840
ATGCCAATAA AAAGTTTTAC TGGCGTGAAA AAGCGGATTA TTTTGTCTAG TAATGACTAA	6900
GTCAGGCACT TCTTGAATAT GATTAACAAA TTTCAAAGGA ACATCGCAAT AATTATTGAT	6960
TTCCTTCGCT AACCAATTTT CTTGTATCTG ATTAATAGAT GAATGAAGTA AAATTTTAAAC	7020
AGGAATTTGT TTATTTAATA AAACACTCCG AATAATTAAA AAAATGGTTT CTTTAAATAA	7080
CGCATCTAGT GGATGTTGTT TGTCAAAGG CAAACGCTTT AAGAAATAGT TGGTTGTTCTG	7140
TGAAACATTT GGGTATTGTT TTTCATACAA AGAGACATTA GCAACCTCAT AGCTCAATTC	7200
CTCTTCCACA GGGAAATGAA TCAGTCGTTT ACTGGCATTG GATAAATTAA CCAACAAAAA	7260
GACCACTTCT ATTGCGCTTA AATCCAACGT TAATTGCGTA CATTTTTCAT AAATCCAAGT	7320
AGCGTGTGGA CTGTTATCCA TTCCCATGAT CGCTAAGTTG GCAAAGTTGA TGGTTTCCAT	7380
TTGGACCAGA GAGTACATCG TTAATGTGCT GAAGAAATCA TAATAAAAAT GAATCTCAGC	7440
TTGCCAAGTG TCAGGCTCTT TTTGTTTCATC AATAAAACAA TCATTCACTA ATAGTTGGAA	7500
TTCTGAATAA GGAAAGGCCA AATTAATAGC AGCAACTTCT GGATAATTGG TGATTGTAAA	7560
ACCGGATTGA TAACGGAGTA AACCTATGGA CATAATTAAT TTTAATTTGC GTAAATCAGT	7620
ATATGTAAAG TAAGGGTAAC GGTTTTTGAT GATTTCAATG AAGTTCTGCA ATTGTGCTTG	7680
ATTAGTCTTA GATAGAAATG GTAGTAGAGA AGTTCCTGAA ATGAAGAATA AGGAATAAAA	7740
GAAACTGCGA AAATCTTTCT CGTTACCACA AATTTTTTCT CGTCTTTTTA AGTTTAAATA	7800
ACAATTAAAT CTAGCTAGAT GATTATTTAA ACGGTCCTG ACTTGTTTAA AAATAGAAAA	7860
GTTAATAAAG TGCTCTTCAC AAAAATCTTT GGCTGAAGTG TAGCGTTTTA AAAGTAAATC	7920
CAGGGCTAAC ATATAGGTTA GTGACCCGCG GACAACAGCT TCTTTCATTA ACTCAATCGG	7980
GATATTTTCA TCGGCGACTA ATTCTATGGA CTGTTGATTA TTAATAATAC GAATATGTAC	8040
ATTTTCATGT GTATTGACTC GAGAAATTAA CAAAGTTAAT TGATTCAATA AATTGTTGAT	8100
ATTGTAATGG GAAAATTCAT ATTTTCGTCG GAGTACTTGC ATGGGGACTT TTTCTTTACT	8160
GTCATACAAG TATTCAATTA GGGAAAGTAA CTGAATTTCC TGTTTGTTTA ATAAGTTATA	8220
CATAATAACC ACCTTTCTAA TTGGTTATGA TTGTTTGATT TTTCAGATAA TACAGATAGG	8280
GCATTCTTTT CTATATGACA TTGACGATTG CCAGTGAAAG AACTAGTTGT TCTTATGCTA	8340
CTGTATCTAT TCTTAAAAAT CAAGAAGAAC TGAATGTGTC AAATGGAAGA AAACATGGAC	8400
AAAAAATGAT AGAGCAAAAA ATTATTAAGA AAAACAACCA GTTTATATAA AACATTTAAA	8460
TAACAAAAAT GATAATTTGA GTCTATTGAT TTTTATAGCTT AAATAGCTTA TACAAATAGG	8520
TTTTTTTGTC TGTTAGTTT TATGAAAAA AATTAAAAA AACGGTATAG TGGAAGAAGG	8580
AATAGTAAAA ATAAGAAAGG ATAAGGACGT GTTTTCATAA AGATGAAAAG GGTTTCTGGG	8640
ATTTTCATCA CTTTTTCTAT TTGGAGGAGA AGTTGTTGAT AATTTTGTCT GAAAAATCTC	8700
GTTATACTCT TCTTATTCTT ATAGAAGAAT CTATAGGAAA AAGCAGAAAA TGAATCCGCA	8760

TTCGAAGTAA GAGTGC GTTT TGAGGAGATA AAAAGAAAAG GTGTGTTGGC TGTCTGTTTT	8820
AGCAAGGAAG AAAGACAACA GAAGAGCACT CACTTATTTG GTTTAAAATA AAAAATTGGT	8880
ACGAAGTGAA CGTTCTCTTC TATGTGTCGT TAGTAGAGGA AGGATGAAAG AAATGAGAAA	8940
GAATGGTCCA ATGGTAAACC GTTGGCTCTA CGGGTTGATG TGTTTGTTAC TTGTTCTAAA	9000
TTATGGCACA CCACTCATGG CTTTGGCGGA AGAGGTTAAC AGCGATGGCC AGTTAACGTT	9060
AGGAGAAGTG AAGCAAACCA GCCAGCAAGA AATGACCTTA GCGCTTCAAG GAAAAGCACA	9120
ACCAGTAACA CAAGAGGTTG TAGTGCAATTA TAGTGCCAAT GTGTCAATCA AAGCTGCACA	9180
TTGGGCAGCG CCCAATAATA CGCGCAAGAT TCAAGTGGAT GACCAGAAGA AACAGATTCA	9240
AATTGAATTG AATCAGCAAG CGTTAgcAGA TACGTTAGTC TTAACGTTGA ACCCTACAGC	9300
TACAGAAGAT GTGACGTTTT CTTATGGACA ACAGCAACGA GCGTTGACGT TAAAGACTGG	9360
TACTGATCCG ACAGAATCAA CGGCAATCAC GAGTTCGCCA GCCGCATCAG CGAATGAAGG	9420
TTCAACAGAA GAAGCATCTA CAAACTCCTC TGTTCTCTCGT TCGTCCGAAG AAACTGTCCG	9480
CAGCACGACA AAAGCGATAG AAAGTAAAC AACTGAATCG ACGACTGTCA AACC GCGCT	9540
AGCAGGACCA ACAGATATCA GTGATTATTT TACAGGTGAT GAAACAACGA TTATCGATAA	9600
TTTTGAAGAT CCGATTTATT TAAATCCTGA TGGAAACCA GCAACACCGC CGTATAAAGA	9660
AGATGTGACC ATTCATTGGA ACTTTAACTG GTCGATTCCA GAAGATGTGC GAGAACAAT	9720
GAAAGCAGGC GATTACTTCG AGTTTCAATT ACCTGGCAAT TTGAAACCTA ATAAACCAGG	9780
TTCAGGTGAT TTAGTTGATG CAGAAGGCAA TGTCTATGGA ACCTACACAA TTAGTGAAGA	9840
TGGTACGTT CGTTTTACCT TTAATGAGCG AATCACGTCT GAAAGTGACA TTCACGGGA	9900
CTTTTCTTTA GATACTCATT TGAATGATTC AGATGGGCGG GGCCCAGGAG ATTGGGTGAT	9960
TGATATTCCT ACACAAGAAG ATTTGCCGCC TGTAGTGATT CCAATTGTCC CAGATACCGA	10020
ACAACAAATT GATAAACAAG GCCATTTTGA TCGAACGCCC AATCCTAGTG CGATTACTTG	10080
GACGGTAGAT ATCAATCAAG CGATGAAAGA TCAAACAAAT CCAACTGTGA CGGAAACATG	10140
GCCACAGGG AATACCTTTA AGTCCGTGAA AGTCTATGAG TTAGTGATGA ATCTTGATGG	10200
AACAATTAAA GAAGTGGGTC GCGAACTTAG TCCAGATGAA TATACCGTTG ATAAAAATGG	10260
CAATGTGACG ATTAAAGGTG ACACCAACAA AGCGTATCGT CTTGAGTACC AAACGACGAT	10320
TGACGAGGCG GTTATTCCAG ATGGCGGCGG CGATGTGCCT TTTAAAAATC ACGCGACGTT	10380
AACAAGTGAT AATAATCCAA ATGGGTTAGA TGCTGAAGCA ACTGTTACCG CCACATATGG	10440
CAAAATGTTA GACAAGCGCA ATATAGATTA CGACGAAGCC AATCAAGAAT TCACTTGGGA	10500
AATTAACCTAC AACTATGGTG AACAAACCAT TCCAAAAGAC CAAGCAGTCA TTACAGACAC	10560
AATGGGGGAT AATTTAACGT TTGAACCAGA TTCTTTACAT TTATATTCAG TGACATTTGA	10620
TGACAAAGGA AATGAAGTCG TTGGAGCAGA ACTTGTGGAA GGAAAAGATT ACAAAGTGGT	10680
AATCAACGGA GACGGTTCCT TTGCAATTGA CTTTTTACAT GATGTGACTG GCGCAGTCAA	10740

GATTGATTAT AAAACCAAAG TTGATGGAAT TGTCGAAGGC GATGTTGCCG TGAATAATCG	10800
TGTGGATGTT GGCACGGGTC AGCATTGAGA AGATGATGGC ACAGCCAGTC AACAAAATAT	10860
TATTA AAAAC ACTGGTGCAG TTGATTATCA AAATTCAACG ATTGGTTGGA CGTTAGCTGT	10920
GAATCAAAAT AATTATTTGA TGGAAAATGC CGTGATTACG GATACGTACG AACCAGTTCC	10980
TGGCTTAACT ATGGTACCCA ATTTCGTTGGT TGTCAAAGAT ACAACCACTG GTGCTCAGTT	11040
GACGTTAGGC AAGGATTTC A TGGTAGAAAT AACTCGTAAT GCAGATGGTG AAACAGGCTT	11100
TAAGGTAAGT TTTATAGGGG CGTATGCCAA AACAAGTGAT GCCTTCCACA TAACTTATAC	11160
TACCTTTTTT C GATGTTACCG AGTTAGACGC TAACAATCCT GCGTTGGACC ATTATCGAAA	11220
TACCGCTGCC ATTGATTGGA CGGATGAAGC AGGAAACAAT CATCATTCAG AAGATAGTAA	11280
ACCGTTTAAA CCTTTACCTG CTTTTGATTT AAATGCGCAA AAAAGCGGTG TTTACAATGC	11340
CGTCACCAA GAAATCACTT GGACGATTGC GGTTAATTTA AGTAATAATC GTTTAGTCGA	11400
CGCCTTTTTG ACGGATCCAA TTTTAACCAA TCAAACCTAT TTGGCTGGGA GCTTGAAAGT	11460
CTATGAAGGC AATACAAAGC CAGATGGTTC GGTGAAAAA GTGAAACCAA CGCAACCGTT	11520
GACGGATATC ACAATGGAAG AACCAAGCGA GAAAAACCAA AATACTGGC GTGTTGATTT	11580
TCCTAATGAT AGTCGTACGT ATGTGATTGA ATTTAAGACG TCTGTTGATG AAAAAGTTAT	11640
CGAAGGTTCT GCTAGTTATG ACAATACCGC ATCTTATACA AACCAAGGTT CTTCACGTGA	11700
TGTGACAGGA AAAGTTTCTA TTCAACATGG TGGCGAATCA GTGAAAAAAG GTGGCGAATA	11760
CCACAAAGAT GATCCAGATC ATGTGTACTG GCATGTAATG ATCAATGGCG CCCAATCGGT	11820
TTTAGACGAT GTGGTTATTA CTGATACACC CTCACCAAAC CAAGTGCTAG ATCCCGAGTC	11880
ATTGGTGATT TACGGTACCA ACGTAACAGA AGACGGAAC ATTACGCCAG ATAAATCTGT	11940
TATTTTAGAA GAAGGAAAAG ATTACACACT GGaAGTTACC ACCGATAATG aAACAGGACA	12000
ACAAAAAATT GTCGTTAAAA TGGCCCATAT TGAAGCACCT TATTATATGG AATATCGTAG	12060
TTTAGTGACT TCTTCAGCGG CGGGGAGTAC AGACACGGTA TCCAACCAAG TGTCATTAC	12120
TGGAAATGGT TCAGAAGTCG TTCATGGGGA TGACAATGGC GATGTGGTCG TTGACATTGA	12180
TCACAGTGGC GGGCATGCCA CAGGGACTAA AGGCAAAATT CAGCTGAAGA AACAGCCAT	12240
GGATGAGACG ACTATTTTAG CAGGCGCCCA TTTCCAAATT TGGGACCAAG CTAAAACACA	12300
AGTCCTACGT GAAGGTACAG TAGATGCCAC CGGGGTTATC ACATTTGGTG GGTTGCCACA	12360
AGGGCAATAC ATTTTGGTGG AGACAAAAGC ACCAGAAGGC TATACAGTTT CGGACGAATT	12420
AGCTAAAGGC CGAGTCATTA CTATTGATGA AGAACTTCA GCCGAAGGAG CACAACCAAC	12480
CATTATTAAA AACGATGTCA ATAAAGTATT TTTAGAAAA ATGGATGAGA AGGGTAAAAA	12540
GTTAGTCAAT GCTCGCTTTA AATTAGAGCA TGCCGTAACC ACGCCGTTTA CTCATTGGGA	12600
AGAAGTTCCC CTTGCGCCCG ATCGAACCAA CGCGAATGGC CAGTTAGAGG TGGATAGTTT	12660
AAAACCAGGG CTTTATCAGT TCACAGAAAT CGAAGCACCG ACAGGCTATC TTTTAGACAC	12720

GACCCCAAAA CGATTCATCG TGACACAAAA TACGAGCGGA CAAATTCGTG ATGTTTCATGT	12780
CAAAATGCTT AATTACCAAG GTTCTGCTGA ACTAATTAAA AAAGACCAAG CAGGCAATCC	12840
ATTAGCAGGT GCTGAATTTT CAGTCCTTGA CACCACAGGA CAAGCAGTTC GAGAACAATT	12900
AGTTTCGGAT GCAAACGGAA AAGTCACAGT GACGGATTTA GCCCCAGGAA AATATCAATT	12960
TGTGGAAACC AAAGCGCCAG CAGGGTACCT TTTAAACACT GAACCAAGTG CTTTCACGAT	13020
TGCAGCAAGC GATCGGGGCA AACCAGCAAC AGTTATAGCA ACGGCTAACT TTGTTAATA	13080
TCAAGGCACG GCTAAATTAA TCAAAAAAGA TGTGAATGGA CACTTATTAA GTGGTGCGAC	13140
ATTTAAAGTG CTTGATGCGA AGGGAGAAAC GATTCAAACA GGCTTGACGA CAAATAATCA	13200
AGGGGAAATT GTTGCAGAGC ACTTAGCCCC AGGAAAATAT CGCTTTGTAG AAACCAAAGC	13260
GCCAACAGGC TATTTATTAA ATACCACGCC AGTCCCATTT GAAATTGCTG AGAAAAATGC	13320
TGGTAAACCA GCGGTCGTGG TTGCTAGTGA CAACTTTGTG AGTTACAAAG GGGCTTTCCA	13380
AATCGTGAAA ACGAATAGCG CAGACCAACC ATTAGCAGGT GCTGTTTTTG AATTATATGA	13440
TCACAATAAA CAATCATTAG GGATTACAGC AACGAGTGGC AAAGATGGcA AAATTATCTT	13500
TAGAGACTTG GCGCCAGGTA CCTATTATTA CAAAGAAATC AAAGCACCAA AATTACCAGA	13560
TGGCGCAGAT TATATTATTT ATCCTGAATT AGTAAAAGTA GAAATTCGTG GTGATTTCAA	13620
AGGTGATCCG GAGATTTTCC AATTAGGGGC CTTCGCCAAT TTCAAAGGAC GCGCCGTCTT	13680
TAAGAAAATT GATGCCAATG CGAACCCACT TCCAGGAACG ATTTTTAAAT TGTATCGAAT	13740
CGAAAACGGG GAAAAAATCT TTGAAAGAGA AGTAACTGCT GAAAAAGATG GTTCATTGGC	13800
TATGGAGGAT TTAGGTGCTG GTAGCTATGA ATTAGATGAA CTGGATGCAA CGGATGGCTA	13860
TATCGTCAAT AAACAACCCA TTTATTTTGT AGTGAAGAAG AATTCAAATG ATAAACAACC	13920
ACTAGATGAG TTAGAGTTTG TAAATTATCA AGCAGAAGTA ATGGGACGTA AAGTCAACGA	13980
GCAAGGTCAA ACCTTAGCGG GTGCAGTTTT TGCAATTTAC AATGCCGATG AGCAGAATCA	14040
GCCCCAAGGT TCACCGATAA CATTCTTGAA TCGTGACAGGA GAAAAAGTTT CTGAAATAAC	14100
AACGGATAAG ACTGGCGAAA TTTACGCTAA AGGGCTAAAT GAAGGGCATT ACGTTTTAGT	14160
GGAAACGAAA GCACCAACAG GCTATCTGTT AGACACAACG CTACATCCAT TTGATGTAAC	14220
CGCCCAATTA GGAAAAGAGC AGCCAATTGC TTTAGGCGAT CTTATCAATT ATCAAGGAAC	14280
TGCTCAATTA ACCAAAGAAA ACGAAACAGG TGAAGCATTG GCAGGTGCGG TGTTTAAGGT	14340
CATTGATGAA ACAGGGCAAA CCGTAGATGG ACAAACCAAT CTGATGTCTG ACAAGCAAGG	14400
CAAAGTCATT GCGAAAAaCT TAGCACCGGG AACGTATCGT TTTGTGGAGA CaCAAGCGCC	14460
AACTAGCTAT CTTCTTAATG AAACGCCAAG CGCAAGCTTT ACGATTGCCA AAGACAACCA	14520
AGGCAAACCA GCCACTGTGG TACTTAAAGC ACCTTTTATT AATTACCAAG GTGCTGCCAA	14580
GCTGGTGAAA ATTGATCAGC AAAAGAATGC CTTAGCAGGT GCTGAATTTA AAGTGACAGA	14640
TGCAGAGACA GGGCAAACCTG TCGCTCGTTC ATTACGTTCT GACAACCAAG GGTTAGTTCA	14700

AGTGAACCAC	TTACAACCAG	GAAAATATAC	CTTTGTGGAA	AGAAAAGCAC	CGGATGGTTA	14760
CCAACGTCT	AAGCAAGCTG	TCGCATTAC	TATTGCGGCA	ACAGCGAAAG	ACAAACCTGA	14820
ACTCGTGAAT	GCGGGCACGT	TTGTTAACGA	GAAACAACCT	GTATCCAAAA	AAACAAAACC	14880
AAATCAGCCA	ACAACGAAAC	AAGCAGCTAG	AGAGACAGGT	TGGCTTGGTT	TACCGAAAAC	14940
CAACACACAA	GTCAATTACT	TCTTTGTCTT	TATCGGCCTC	ATGTTGGTCG	GTTTGGCAAG	15000
TTGGCTCTTC	TATAAAAAGA	GCAAGAAATA	AGCAAAAGCT	ATAAAAAAAC	AGAAAGrGCC	15060
TGGAACAAAA	ATCACTTTGG	ATTTTTGTTC	CAGGCyCAAA	ACCTTATAAA	CGGCGGTAAC	15120
AGAACCAACT	CCTTCGGAAA	TAAGCCGAAA	TTCTCCAAAA	ATTAAAGAGC	AATTTTCGGA	15180
AATTTCTTCT	TATTTCTCGG	AGCTAAACGG	TTCTGTACCG	ACCTCTTTTT	TCATCTTAAG	15240
CTAATCAACT	GAACAAAAGC	AGCTCCTTCT	TATTGACATT	CTCAGGTAAG	CCTTCATCTC	15300
AGTTTACCCT	TTGAACAGAA	CTTACTGGGT	TCTTGGAGAT	AACCGCTTTT	GCCTCAGCCT	15360
TTGTTTTTCG	ATGCTATCTT	TAGGTATTTT	GGCTAAAATA	TTGTACACTT	TACATGTGCA	15420
AGAAATTTAG	TTAGAGATAT	GAGGATAGAA	TGATGAAAAA	GATGATGGTT	TTATACAACG	15480
AAACTTCAGG	AAGTAGTGAA	AGTAAAGAGA	TTGCAGAACG	CTTTAAAAAA	GCTGCAGAAG	15540
CGAGAGGGGA	AGCGGTTATT	TTGCAACCTT	CCAATCCAGA	CATTGATCCT	GAAGAAATGC	15600
GGAAAAATGC	CAAAGAAAAT	CAAGTGGGTG	TCTTGGTCGT	CATTGGTGGT	GACGGGACCA	15660
TTCATCATGC	CGTTCAAAT	TTTAAAGACA	CCATTCGTGA	CTATCAAATA	GGGATTATTC	15720
CAGGTGGAAC	AGTCAATAAT	TTGCACGTG	TGTTAAGTAT	TCCTTTGAAA	GAAGAAGACG	15780
CCTTTGAGAC	CATTTTGGCA	GGACAGACCA	CACCAAGTTGA	CTTTGGGATG	GTAAATCAAG	15840
ATGTGATGAT	TTCGACCTTA	ACGATTGGAC	TATTAGCAGA	TACAGCAGCA	AATGTGACCC	15900
AACAGGAGAA	ACAAAAATAT	GGTCCTTTAG	CCTTTACAAA	GCAATTTTTT	CGCTTATTAA	15960
TGAAAAAGAA	AAAATACAAA	CTAAAAATTG	ATGGCGATGA	AAGACGTTGG	CATGGCAAAG	16020
CCCAATTATT	AACGATGACC	ATGACGAATT	CTGCTGGTGG	CTTTACCAAT	TTCGATGCCA	16080
ATGctACGCC	AGATGACGGC	GAAGTCCACA	TTATCATTTT	ACCTAAGTTA	GTTTTTTTACA	16140
AGTTTGTCTA	TTATTTACCT	AAAATTATCA	GAGGACAATT	AAATGAAATT	CCTGGTGTCTG	16200
TGTATTTCTC	TGGGAGTCAA	TTTAAAATTA	GTGGCGAAAA	AGACAAGAAA	GTCCAAACAC	16260
GAACAGATGG	CGACCCTACT	GATGATTTAC	CGATTAAAGT	AACGGTGGAA	GCTGGGGGAC	16320
TCAACGTGTT	TGTTCTTGAA	ACGTCCACAT	CTACGAAAGA	ATAAAAGAAA	CAACTGTTAT	16380
TTTTTCGTGG	AATAGATTTT	TCATTTTGTG	TACAATAGAA	CAGAAGAGAG	GCAGTTGAGT	16440
GAACAGTGAA	ATGGGATGAA	TATGTTGATA	TATTCGATAC	AGATGTCGTT	TCTAAAATTA	16500
TTATTTTAAG	AAACTTGCCA	ATTAATGAAG	GCTGGATTGA	CATCAATGAG	TTAGCAGAAG	16560
AATTACAATT	AAACAAAAAA	TCAGTCAAAA	AATACGTAGA	ACTTCTAAAA	GATGACATTA	16620
CGTATTATGT	TGTGGATAAT	GATGCTTCTT	TGGCTTTTGA	AAAGGGGCAT	GGCTATCGCT	16680

TGAACTTGGC	GTCTAGCCAA	GCGTTCCAGA	AACTTTATAC	AGCAATTTTT	CGGGATAGTT	16740
TAACTTATCA	AGTAATGGAA	GATTTATTCT	GGGGTCGATT	TAATGGAATT	GTGCCAGCGA	16800
CAGCAAAGTA	TCATACTAGT	GATGCCACGA	TTCGGCGCTT	GCTCAAAGGG	TTTACAGAAA	16860
TTTTAGCACC	ATTGGGTTTA	GTAATTAATA	AAAGTACTTG	GAAAATTGAA	GGCAATGAAG	16920
CGTATTTGCG	CCATTTTGCT	TACACTATTT	TTATGGACTT	AATCCGTGAT	ACTTGGCCTT	16980
TTGACTATAT	ACAAGAAACG	GAAGTCGCAC	AAACCGTCCA	AAAAATGATG	GCTTTTTTCA	17040
ACGTAGAAGT	TGCGGAAACG	GTGACACGCC	GAGTTAAATA	TATGGTCGCT	ATCTCAAAAG	17100
TCCGTAGTGC	GCAAGGCGCC	GAATATCGAC	CTTCAGCCGA	GCAAGAACT	TATTTAGAAG	17160
GTAATAAACA	TTACCAAGCC	TTTATAGAAG	AGATGACAGC	TACAACCCTT	TTCTTAACA	17220
AAAACGATTT	AACTTTTG TG	TACTTTTCT	TGTTAGCCAA	TGACCAATTT	TATCAAGATG	17280
AAGCCACTGC	ACAAGCAATT	TTAGCGCATT	TTGATGAGAC	TGAAGCACCT	GTTTACATGT	17340
TGACTCATTT	AGTCCAACAT	TTCTTATTAG	AAGAAATACA	CATGACGCCG	CAATTGGAGA	17400
AAATTCGTCC	GCAAATTTTT	TATTATTTGT	TTGCTACACA	TTTATTTGCC	GAACTTTACT	17460
ATGTCCAAAA	TGTGAAAATG	TACAATGTCT	TTTGAATATA	AATCAAAGAA	AAATATCCGG	17520
TTCTTTTAGC	AGAGTTAATT	CAAGGTTTAA	ATCAACTTTA	CGAACAAACA	GGCAATGAAT	17580
TATTTACAAA	TAAAGAATTT	TTGGCTTTGC	GTTATTCTAG	TGTTTTGGCG	CTTTCCAATG	17640
ATTTAATTCA	TCGCGAACAA	AAAATTGTTT	TAGGGTTAAA	ATCAGGCTTG	CCAGTACTGG	17700
AAGAAGAACG	GGTGAAACTT	TCGATTGAAA	ATAATTTTCG	TCGGTTTTAT	GCATTGGAAG	17760
TTTTACGATA	CAGTGAAGCA	AGCAAAGAAG	AATGGCAAAA	TATTGATTTA	CTATTAGTGA	17820
CTTCTTTGGC	AGATGTGGAA	ACGGGCCATC	CAGAGATCTT	TTTATTTTCA	AGTGAGTTAA	17880
CGTTTGAGCA	ATATAATCGT	TTGAACCAAA	AAATAGTAAA	AATTAACGAG	AAAAGAACCA	17940
CGTCACTTTA	ATAAAAAATA	GGCTATTTTG	ATGCACATAT	CTGATAAATT	GTAAAAAAT	18000
TTTTCTAAAA	AGTAAAAAAA	GACAAGAAGA	AAACGTCAAT	AAGTAGGTGT	TGATTACCTA	18060
CTTATTGACG	TTTTCTCGTT	GTTTTTTAAT	AAATACTGCG	TCAATTCCTC	TGAAAATTAT	18120
GTCAAAAAGG	TACTTTTTTT	GCTTTTTTTA	TCGTAAATAA	TAGATGATTA	GCTATTGATA	18180
GCAAAGGAAC	TACTACTTTT	TTTAACAGTC	AGTTAAAAAA	TATTGAGGTT	TCTCACAAGA	18240
ATTTTCTAT	ATAATAGAAA	GAGCAAGGGA	AGGAAGGCCG	GATAAGATGA	AAAAACAGAT	18300
GGAACAAATT	GAAGAAATCT	TAAGTTGTGA	AGAAAATAGT	GCAGGCATCC	GATTAAAGGA	18360
ATTGGTTGAA	GCTTTAGAAT	TGGAAGTCAC	AAATCAAAAC	TTGTTAAAGG	TAACAAGTAT	18420
CTTACATATG	AATCCAAAAT	TCAAAAAAAT	TTATGCATAC	GAAGACAGTC	GTGTTATTAC	18480
GCTGTATCAA	CTTTTACAAA	ATAAGCCGTT	GGAAGTAACG	GAATGAAAAG	AGGGAAACAC	18540
ATGAAGAAAA	CAATGCTCAT	TCTTCTATTT	GGTTTATTGA	CAGTGGTTGG	TCTGCCAATG	18600
GTGACCGAAG	CGGTTGAACC	AGTTAACGCA	ACAGATACAA	CGATTTTTGG	TGCGCAAGCG	18660

ATGGTGCCTA ATAGCACAGA AGATCAAACA GATAAGTTAC AGACGTTGCT GAGTCAGACA	18720
GCAAAGGAAG GACGGGCATT ATTTTACCA CAGGGAAGTT ATGCTCTGTC AAAAGATATT	18780
GCGATTTCTT CAAATTATCA ATTAATTGGC GATACAACAG GAGCCACTAT TTTACATAAT	18840
GCTACAGGTG CACCGATTCA GCTAACAGAC ACAACGTACG GGACGAAAAC CAATGTACGA	18900
CTTCAAATA TTGCTTTTGA CGGCATCAAT GTGACCTTAA AATTGACGAA TCAATTAACG	18960
CTAGCAAACA ATATCTTTTA TAATCCATTA AAAGGATTCG TGGTCAATTT AAATGCAGAC	19020
ATTGGCGTTA AGATTTTCAGG CAATATTTTC ATGCGAGACA CGGCACATAT GCAATCGGGA	19080
GGTGATTTTA ACCGAGCGAT TTACATTGGT GGTATTCAA CGCCAAGTCG CTTCCAGTAT	19140
ATGTCAGATG TAGACATCGT GGACAATCTA TTTGGTTTAA AAGTGACAGA ATTAGATGCG	19200
ATTAAAAGTA CCAGTCGTTT AGACTTAGCA GCGACTATTA CACGTTTGCA AACAGCCATT	19260
GAAGCCGGTG CTATTTTCAGT ACCTAATGAA CAAAATTATC TATCCACCGG GGTAAATCT	19320
TTTAATATGC TGAAGGATGT AACTGTTCAA CATAATTTT TCTATTCACC CTATGATAAT	19380
GAAAACTTGA ATGGTCTAGT GGGAGATCAT GCAATTTATT TCCGTGGCGC CAAAACATC	19440
ACGGTAGTAG GGAATCATT ACGCGGACTA CAAAATGGTC CAGCTGGCGG ATTTAAGTTT	19500
AAATCAGGCC GTAATATTAC GATTATGAAT AATTATTTGC GAAATACAGG TCTCATTATG	19560
TATGGAACAC CAGAGATCGG TTTGGCAGAA ACACAAGCAG AGGGGGCTAT TTCAGAACTC	19620
TCTAACTGGT TAGTGGCTAA TAACATTTTT GATTGGAAGT ATTGGGACAA TCAATATGCA	19680
ATTGGGATGG AATACAATCG ACATACAGGA AATAACAATG TGTTCATGG CGTATTTATC	19740
AACAATCAGT TTGTAACTA CCATAATATT CCGCAAAACC GACGTCGGGA ATTGTTAATC	19800
GCTAGTGGCG GTGGGTTCCG ACCAGAAACG AGTTTTGTGA AAGACAATAC TCGAGATGAT	19860
GGCCTTAAA ACGGCCAATT ATTGGTAGAA AATTGGACAG AGGCGGACTA TCGCTTAATG	19920
CCAGCAACAT GGAATCATT AATTTACCA ACACCTTATG AGCAATATAA AAATACGCCA	19980
ATTCCGGTTA GAAATACTTT AGCAACCCCT GTGGCTACAA CGATTGTGCA AGGACAATCA	20040
ATCGATCCGC AGCAACTGGT TGCTAATACA AATGACGCGG ATGAAGCAGT CCCAGCCGCT	20100
AAAATCGTCA ATCCAGAGGT CTAAATGAA ATCGGCCAAC AAAAAGTGAC TGTTCAATTA	20160
ACTTATGAAA CAGGGTCATT GGTAACAGTG AACGTACCGG TGACGGTAGA GGCCCTGCA	20220
AAAAAGCTTG ATTTGTCCCA GCTACAAACC GTTTATGCTT CGATTGGTGA AGCAAACCAA	20280
TATACAGTTT ACTCTTGGCA GCTATTCCT GCCATTGGGC CTAAAACAAT TGTTCTTCC	20340
TATTATCAGC AGGCGACACA ATTGTTAGCA GAAGGTCAGG AGAGTCAGGA CAAAACCCAA	20400
GAGCAAGTGG ATCAGTTAAC AAGTAATTTG CAATCTGCAA TGAAAGTATT AGTCAAAAAA	20460
GCCGATATCA CGCTAGAGCG TACAGAAGCA GAAAATGAAC TGGCAAGCGT TCATAAATTA	20520
GACGAATCTG TATATACCAA AGACTCTTGG CAAGCCATGC AAGAAGCCTT AATAGATACA	20580
ACAACTGGCG AAGGAAGCTC TAAGCAGTTA CAACAACCTGC TTGCTTGGTC AGATGAAGAA	20640



TTATTAGAGC	CAACACTTGG	TGGCTTTAAA	ACGCCTGCAG	ATGCACAAAa	GCGTATTAAT	20700
CAATTAACAC	AAACCATAAA	AACGGCGTTA	CTTCTGTTAG	TTGAAAAATC	AACAGAGACA	20760
ACTAGTAACA	CAAGCGAAAG	CAGCACCTCA	AGCACAACCA	GTGAAACCAG	TAATACAAGC	20820
GAAAGCAGCA	CCTCAAGTAC	AACAAGTGAA	ACAAGCAATA	CAAATGAAAG	CAATACACCA	20880
AGTACAACCA	GTAAAACCAG	TAATACAAGC	GAAAGCAGCA	CCTCAAGCAC	GACAAGTGAA	20940
ACAAGCAATA	CAAATGAAAG	CAACACACCA	AGTACAACCTA	GCGAAACCAG	TAATACAAGC	21000
GAAAGTAGTA	CCTCAAGCAC	GACAaGTGAA	AGCAGTAGTA	CCAGCGAAAG	CAGCACGCCA	21060
aGCACGACAA	GTGAGAGCAG	TAGTACCAGC	GAAAGCAGCA	CyTCAAGCAC	GACAAGTGAA	21120
ACCAGTAATA	CAAGCGAAAG	CAGCACGCCA	AGCACGACAA	GTGAGAGCAG	TAGTACGAGT	21180
GAAAGCAACA	CGCCAAGCAC	AACCAGCGAA	ACCAGTAGTA	CCAGCGAAAG	TAGTACCTCA	21240
AGTACGACAA	GTGAAACAAG	CAATACAAAT	GAAAGCAACA	CACCAAGTAC	AACCAGTAAA	21300
ACAAGTAGTA	CCAGCGAAAG	TAGTGCTTCA	AGTACAACCA	GCGCAACCAA	CAATACAAGT	21360
GAGAGTAGCA	CCCCAAGCAC	TGTAAACGAA	AGTAGCCAAA	GCAAAGGCCA	GAATTCTGTC	21420
ATTTACGCGG	TAGAGTCTAA	TCAAGATCCT	AATGACGCTC	AAAGTAATTC	AAAGCCCTCT	21480
GCCAAGGCTA	GTCAAACAAA	GGAATCAGTA	GCAGAAAACC	AAGCAACTAA	ACAGATACAG	21540
ACAAATCAAG	AATCGTCAGG	AACAGTCAAA	AAAGCGGATA	ATACGACCAA	AATTGCCAAG	21600
AAAAAATTCC	CCAAAACAGG	GGAACAATCT	TCAGCAGTTG	GGAGCTTCCT	AGGCCTGAGC	21660
TTTTTATCGT	TAGCCATAGC	AACCTATTGT	TTTAAAGTGA	AAAGATAACT	CTTTAAAAAC	21720
GGATAAGTAA	ACATCCTGAA	ACAAGAAATA	GGACTTGTTT	CAGGATGTTT	TCATATAAAA	21780
TTAAAAGAAC	TACAGTTTTT	AGTGAAAAAA	CACACAAAAG	TACTTGAAAT	TCTGTGATAT	21840
AAGGGTAAAG	TAAAGGAGGA	CAAAAAACAA	AGGAAGTGAA	ATGTTTGTAT	ACACCAGATG	21900
AAATTTTAGG	TATTAGTATT	GAAATGGTC	AGAAAAAAT	TCAAAGCCT	TTGGTAGCAA	21960
AACTCATTTT	AGGATTTATC	GGCGGTGCCA	TGATTTCCCT	CGGCTATTTG	GCTTATGTGC	22020
GTGTTTCCGC	CTCTATTCCA	GCAGATTTAG	CCAGTGTTCA	AGCTTTAGTG	GGAGCTGCGG	22080
TTTTCCCAAT	TGGTCTAATT	GTTATTTTAA	TGGCTGGTGG	TGAGTTAATC	ACAGGAAATA	22140
TGATGGCGGT	GAGTACGGCT	TGGTTTGCTA	AAAAGGTCTC	TTTCAGAGAA	TTACTCGTGA	22200
ACTGGGTCAC	AATTACATTG	GCAAATATGG	TTGGCGCCCT	ATTTGTTGCC	TATTTTTTTG	22260
GCCATTATTT	AGGTTTAACC	AGTAGCGGCG	CCTATTTAAA	TCAGGTTCTT	TCTCTTGCAG	22320
AACATAAAGT	TGCAGCAACT	TGGCTTCAAG	CAATTATTTT	AGGTATTGGC	TGTAAGTGGT	22380
TCGTAGGTCT	GGCCTTATGG	CTTTGTTATG	GTGCGAAAGA	TGCAGCGGGG	AAACTATTAG	22440
GCATCTGGTT	TCCAGTAATG	ATTTTGTGTG	CAATTGGCTT	TCAGCATAGT	GTGGCGAATG	22500
CATTCATCAT	TCCAGCCGCA	ATTTTGAAG	GCGGCGTTCA	GTGGGGAGAC	TTTTTCCGTA	22560
ATTTTATTTT	TGTTTATTTA	GGTAATATCA	TTGGTGGTGC	CATTTTGTGT	TCTGGCTTTT	22620

ACTATTTAAG CTATCGCAAA CATTAAATCC ATTTAGCAGA TAAAAACCTA CACAAAGGCT	22680
AAATGCCTTT GTGCAGGTTT TTTCTTTTTA AACAAAAAGG AAAACGGTTG AAAAATTATT	22740
GCTAGATTTT TAAAACAATG TTACATTAAT GTTACAAAAA GATTTCTGGA GGTAACGAAA	22800
AATGAAAAAA ATTGCTTTGT TCAGTATGTT AACGTTCACT GTATTGTCTT TAAGTCTAGC	22860
AGGATGTGGA AACAAAAAAA CAGCAaGCAC AAATGATTCT AAGCCAAAGC AAGAAACAAA	22920
GAAAGCCACG CAGAAATCCT CTAGCCAACA AGAAATGAAA AGTAGTCATT CGTCTGTCAC	22980
GGGTCAAAAT TCTAATGTGA CAGGGGAAAA TCCGTCAGAA AATGCCACGC AGCCTTCTGC	23040
AGGAACTGAT GAAACGAATG AAGTCCCTCA AAACCAAGCA CCTGATACAA ACATTACAAT	23100
TACCAATGTT GTTTTCAATC CTGAAAGAAA TGAAATTAAT GGTACTACAT TACCTAATGC	23160
AACCATTACA GCAACGGTAG TCGGTGATGC TTCTGCACAA GCAGGTGTTT TTTATGCGGA	23220
TGCCAATGGC AATTTTACAG TAATTAGTCC CAGAGCGGGA GCGACTACTC AATTAATCGC	23280
AACCGTTGAT CAACGGAATA GTGCACCTGT CCAAATTGAT ATTCCAAGTT CAGGACAAGA	23340
AGCAGCGCTT TCTTTTAGCA ATATTACGAT TGATCCGAAA CAAGGGACAA TTTCTGGTAA	23400
AACAGCACCG AATGCAACTA TTTTAGTGTC ACGTGCAGAT GATGCGCGGG TGATTTTAGC	23460
AAGTTTTACT GCGGATGCCC AAGGGAATTT CACAGCCAGT AATTTAGTTC CCGGCACAAA	23520
AAATCGCTTA GATGTTACGT TAAATGGAGA AATAGGGACA CCTTACTTGT TTGATTTACC	23580
AAATTAAAAT AGAAGAGAAG TGATCAAAGC CACCAACGCT AGTTAGCGTT GGTGGCTTTG	23640
ATGGTTTAAG CTTCTTCTTT TTTATGATTG CGGATATGTA AATAGATGGT ATTTTTTGAA	23700
ATACCTAATC GATCGGCAAT CTTTATGACA GAATCTTTGA GATTAAAAAT TCCTTGATCA	23760
TATAAACGAA TAACAATTTT TTTATTTTGA TTCAAATGTG ATACATTTGG ATCGCTTTCT	23820
ACTGCTTGTT TTACTTTTGT CAGTGAAGTA GTGATTAATT CGTCAATATT GTCAGAAAAA	23880
TGTTCAGAAT TTTTAAGCTC TTGGGACATA TCATTTTGTT GTCCTGAAGG TAATAAGTCT	23940
TGCAAGTAAT CAATTAATGG CATTTCTAAA TGCATATTGA TACAGATCAA TCCAATAATA	24000
CGATCATTTT CACCGATAAC TGAATGGTG GAAGATTTTA AAATAACCCC TTTTGTATT	24060
TTATTGTAGT AGGGTTTCAC AGCGTAATCA GTTCCGACT CTAATTGACT TAACATTTTT	24120
AACGCCATAT CAGTAATGGG GGCACCTTCT TCGGTTACT GAAATGGCCA TTGATGACTT	24180
TGATAGCCGA ACTTTCCAAA TTTTCTAAAC TATGAAGAAT GACTTCATGG GTACTACCAA	24240
AGAAGTTCCC TAATCCATCA AGCATTAAGG CATGTGAATG TAAGATTTGT TTATCTACTT	24300
CACTTAAACG TAAATATGAC TCCaAGTGAT ATCCTCCCTT AACTGAAAT TAATTTTAGT	24360
GTAATACTTT TTTAATGGAA AAGCAATAAG ATTCTTTACG TtAAAAAAAc nCaTtGACTA	24420
AAcTTTtGTT TAGTCGTATA CyCKGAATGT GGAATGAAAT TTTGTACTT CATTGTACTT	24480
TAGTTCATCG GAATGCTCGC TCATTCTTTC ACTTGTTGaG CaATAAAAAAt TTTaTTCAGG	24540
AGTACCATCC GTTATAAAAA AATATAGTTG GAGGCAAAGA AATGAAAACG TTATCAGGAA	24600

TCTTGATGTT AGTCTTAGTT TTATGTGGCA TGTCCCTATT CTCTTTTAAA GCACCGCGAG	24660
GGAAAAAATC TTTAAGTGCT TTGTCTGGTG CAGCTTGTGC CACCTTTCTA CCAGAGGCGT	24720
TTTTGCGGTA TGCAGTAGGC GCGGTTTTTC ATTTAGATTA TGTGTCGCAA ATTGGTGAAA	24780
CTATGGGTAG TCTAGGTGGT TTAGCAGCTG GAGCTCTTGT GCCGCTTGCT TTCGGGATTA	24840
GTCCAGTTTT TTCAATTATT TTAGGTCTGT CATTGATTAA AGTCAGCTTG TTACCAGCCT	24900
TTATTACCGC GTATCTTTTA TCTTTTGTTG TTGAGCAGAT GCAAAAACGT ATTCCAGAAG	24960
GGATTGACTT ATTAGTCGTT ATTTTAGTCA TTCCAGTCGC TGCTTCGTTA GTTGCTGGTT	25020
TTATTCAACC CGTTGTATTA GGTGTCCTAG AAGTTATTGG GGGCACGATT TTAAGTGCAG	25080
TAGATGGCAA TCCATATGTG ATGGGCGCAA TTTTAGGTGC TATTATCCCG ATTGTTGGTA	25140
TGACACCGTT AAGTTCAATG GTTTTAACTG CCTTGATTGG TTTGACTGGT GTACCAATGG	25200
CTGTCGGTGC GTTGACTTGT TACGGCAGTT CCATTGTCAA TGCGGCGCTA TTTAAAAAGT	25260
TAAAACTAGG CACAGCTTCA ACCCCGTTAG CTGTGGCAAT TGAGCCATTA ACACAAGTCG	25320
ATATCATCAG TTCCAATCCA ATTCCTATTT ACGCAACGAA TTTATTTTCA GGAATGGTTA	25380
GTGGCATTGT AGTGACCTTC TTTGGCTTAA AAGTACCTGT GACAGGAATG GCAACACCGT	25440
GGGCTGGTTT ACTTGTGACA TTAGGTAACA ATGCCATTCA AACGACTTTA CTAGCAGTAG	25500
CGATTATCAC AGTGGTCAGT CTGATGTTTG GATTTTTAGG TGCGCTGGTC TTTAAACGTT	25560
ACAGAATTGA AGCGGTGGAT AATACACCAG AATTTGAAGG aACGACGGAA GTAACCAGTG	25620
AAAAAGAAAA AGAAGCAGTA AAAACCAAAG GAGGAGTTAC GTATGAACAC AGCTATGGCC	25680
ATCAATAAAC CTA AAAAAGT GAACAAAGTA ATGAACTACA AAAGTTGTTT TGACATTATC	25740
GGTCCTATTA TGATTGGTCC TTCTAGTTCG CATACAGCTG GGGCCTTAGC GATTGGCTTA	25800
GCGGCAAGAA AGTTATTTCG CGGAACGCAG AAAAAATTGT GATTAAATAC TATGAATCTT	25860
TTGCAGACAC GCATAAAGGA CATGGGACTG ATTTTGCAAT TATTGGCGGA ATTCTTGGAC	25920
TAGCAGCAGA TGATGCCAAC GTAACACAAT CCATTGAACT TGCGGAAGAA CAAGGAATTG	25980
AGATTCAATT TTTAGAAATG AGTGAAGAAA GTCCAGTGAA GCATGCTAAT ACAGCATGTG	26040
TGACACTCAG TGATGCAACG CACGAAATTC ATTTAACAGG AATTTCTGTA GGTGGCGGCA	26100
CTATCGAGGT GAAATACATT GAATTAGACG GTTTCAATGT CCAACTACAT GGTCCTTTAC	26160
CTATTTTGT AGTCATTAAT CAAGAAGAGC AGGCGATGCA AGCTTTTAAA GATACGTTAC	26220
AAAAAATAA TATTCAAGTG AATGCGGTGT CTCGTTATGT GGAAGGGAAT CAAATTTTAT	26280
TCATCTTTGA CTTAGATTTC GCACCAATTT CCTCGGTGAA AGAACAACCTG TTTTCACTAG	26340
ATGACACATC GAAAATAATC TTATTATGAT AGGAGTTTGA CATATGTATC TTTCGATTGA	26400
AGAATTTATC GATAAAGCTG AAAAAACAGG AAAAAAATC TCTGATTTAA TGATTGAGCA	26460
AGAAATGGAA CGTAGCCAAA AAAGCTATGA AGAAATTTGG ACACAGATGG GTCAAAATTT	26520
GGATGTCATG GATGCTGCTG TCAAACGCTC GCAAGAAGGA GCAGGCGTTT TTTCTCCTAC	26580

TGGGTAAACG	GGCGGTGATG	CGGCTCGTTT	AAAAAATAC	CGCGCAgcTG	GCAAAACGTT	26640
ATCAGGTGAT	TTAATGATGT	CTGCAGTTCA	AGCCGCACTA	GGTACTAATG	AAGTCAATGC	26700
AGCAATGGGC	GTTGTTTTCG	CAACGCCAAC	CGCTGGTGCA	AGTGGCACCT	TACCAGGGGT	26760
TTTAACAGCG	ATTAAAAATA	CCTTGAATTT	AACGCGTGAT	CAACAAATTC	GTTTCTTGTT	26820
TACCAGTTCT	TTATTTGGAA	TGGTCACAGC	CAACAATGCG	ATGATTGCTG	GTGCTGTTGG	26880
TGGATGTCAA	GCCGAAGTAG	GCAGTGCTTC	AGCGATGGCC	GCCGCTGCAG	CTGTTGAAGC	26940
GGCTGGCGGA	ACACCACGCC	AATCTTCTGA	AGCTTTTGCT	ATGGCTTTAG	GGAATTTATT	27000
AGGTTTAGTT	TGTGACCCAG	TGGCTGGGTT	AGTCGAAATT	CCTTGTGTCA	AACGAAATAC	27060
AGTAGGTGCA	GGCAATGCGT	TAATCGCTGC	AGATATGGCT	TTAGCTGGAA	TTGTAAATAA	27120
AATACCAGCA	GATGAAGTGG	TGGAAGCCAT	GAATAAAGTC	GGCCGTCAAT	TACCACGAGA	27180
ATTGAGAGAA	ACGGGTTTAG	GTGGCTTAGC	TGGCACAGCG	ACTGGCCAAC	GAATGAAGAA	27240
TGAAATTTTT	GaAAAAGTAA	AATTCACCAT	TGnTTAAAAT	TTAATCATGA	AGAGGaAtAa	27300
yGCTGGTTGA	AAAATTCGCA	ATGTGAGTGT	CAGAAAGTAA	ACCGGTnGGC	nAGAGTTAnC	27360

## (2) INFORMATION FOR SEQ ID NO: 165:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11427 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

CAGAAACTGA	ACAAAGTGAA	TTCCCTTTTG	CGATTCAChA	AAGTCACATT	TTCAcGcAGG	60
tTTaAatCTt	CCTGAAaaTT	TtAAaGGTaT	TTTTttGCTt	GAaGAaaGTT	ATTATaCTTT	120
AAATGGAAAA	GAACGATTCA	AATCAGATAT	CTTTCTCTTT	GAAGTAAATT	CGTCAAGTAA	180
CGTTAAATTA	AGTTCCATAA	CTGTTCCCTAA	AACTATGCT	AATAAGAAAT	ATGAAGAAAT	240
TGAATCTATT	CCTTTTTCTG	AAATGACCAT	TTCAGAAAAA	TTTGTTCCGA	TTGTTTACAA	300
AAAAGAGAAT	GAAACTTTTA	TTGGTCAATC	TGAAAGTAAA	TTTACAGAGA	AGAATCATTT	360
CATATTAAAA	CAAACGATAT	CTGAAAAAAA	ATTGATAATC	AAAGAAGAGA	TAATGAGAGG	420
AAATAAACGG	ATATTCCGAT	TCGATAAACC	TATCGAATAC	GTAAGAGTTA	AAGAAAATTA	480
AAAAAGCTTT	TAAAAGAATA	TAATTTAACA	GTGAAGTAAT	TTCTTTGTAA	AATAGTAATT	540
ACTTCACTGT	TATCATATTA	ATAAAGATTT	TCATCATTTT	TGAAAAATATC	TTCTCTTTCC	600
AAGAAGAATG	CTAATTCAGA	TGTGTGTGGT	TAAATTGATA	GTCTGCAACA	TTAAACATTT	660
GACTGATTCA	CTTTCCAAAA	TTCTATTATA	CCAGTTAAAA	TCGCTTAAAT	AGAAATACAC	720
TTCTAGCAAC	ATGTTCCATA	GTAAAAGTAG	CACTTGTTAA	AAATAACTAA	TCTGCAACTT	780
ACTATCCTTG	AAAATAAAAT	TCTCAAACAC	TTTACAAACA	ATTATTTTTT	AGGAAGATGG	840
TTTGTTATTC	TTATATTTTCG	AGAAAAATAA	AAATAGCAAA	AGATACATTA	CGATACAACT	900

ATACTATCAA TACAAATTA GATACCTCTT CTGCCCTTTG TTCAGTGATA AAAGCAATTG	960
AGTTAGGAAT AATAAATGTT CACCTTCTAT AGACTTATTT ATTCAGTCAG CTAAACGCTG	1020
TTAGATACAG CTATAGAAAT CTATTATCTG ACATAGTTAG ATAAACAACAT ATCCTTTATA	1080
AAAATTGGTT CCATGTCTTT AGTACAAAA GAGTACTATC ATTATTTAAT TTAAGTAAAT	1140
AACGGATTTT AGAACTCATC GTGATTTCTA CGAGCACTCT TATTCAATTG GATAAGCGTT	1200
ATCTCCTCTT TTTTAAAGTA ATTGTATCTC AGCAACCTAG AATATTGATA GTGCCGATTA	1260
TTTAGACAGC TATCCGGATA ACTTGGAAAT ACTGAACTA GAGTAAAAAA TCTAATAAGG	1320
TGCTTTTAAAT ATATCATCAT TTAAGATCTG AAATCGCATA CCCCAGAAAA TATCTGCTCT	1380
CTTTAGTTCT ACAATTTCTG ATATCCAAAT AGTTCAATTT GAGTTAACGC TGGAAATGGT	1440
GATTCATCTT CTGCTTTGAT CAGCTGTCTC AATGTAGCAG TCGTTACTGT TCGTGGAGAG	1500
AAGGTGAAGT ATTGCGTTGC TATCGCATCA CTCGTAGTGA ATATTTCTTC TGAACCATCG	1560
GAAAAGGCTA TGGTTACTTG TTTCCAATA CTATCATGGG GAAAATCAGC ACGTAATACC	1620
AATCCAAGCT TATCGAGTTC AACTTCTCSG CCAAATCAA TTGTTATTTT AGCATCTTTT	1680
TGCTGATTGA TCCCCAAGA TTGATAGGSG TACGATCCAT GATCTCTGTT GGCCAGAATT	1740
CCATCAATTG CATTTCTAGC AAAAAAGTC GCATCGTTTC GTGTTTCAAC ATTTGCATAA	1800
GCGTGGGGAT AAACATTGGC CACATCTTTT TGATCATGAG GATTGAGCGC CCAGTTACGA	1860
TAAGTAGTCT GTTCTTCAGC AAAAGCCTTT CTGACAAAA AGTAATGCTT CTCGTCAAGA	1920
AAAGCCGCTG GTGGATAGGC TTCTCGCAAT TTTTCCGTCG TAGGAATTTT ATAGACCCAC	1980
TCGCCTTTTG ATAAATACAC TAAAGAGGAT CCTAAAGCAT TGTCTAGTTG AACATAAACT	2040
AATTGTTTTT CTGCGGTTAC AACTTTAAAT TTATCTCCTG CTGAAAAGC ATACTTTTTC	2100
GTTGCCAAAA CTGCTTCATC CACTCCACTG CTGCACATAA CTACTTCATT TTCTTGATTA	2160
ATAATTGATA ATTCTATTTG CATACTTTCT TCGCTCCTTT TACTTTTCTT TTAAAACTT	2220
ACCCGCAATA AATAAAGGA GAGGCGGCTT TAACCGTTTC CTCCCCTCTT TATTCTCTTA	2280
AGGTTGTTTA CCAATATAAG CTAAAATTCC GCCATCAACA TACAGAACAT GACCATTTAC	2340
AAAGTCAGAG GCCTGTGAGG CAAGAAAGAC TGATGGACCT GCAAGATCGA CCGGATCACC	2400
CCAGCGTGCT GCTGGCGTAC GACCGACGAT AAATTGGTCA AATGGATGAC GTTCCCCGTT	2460
TTCTTGAGC TCTCTCAGCG GCGCTGTTTG CGGTGTCGCA ATATAACCAG GACCGATACC	2520
GTTGCACTGG ATATTGAATT GACCATATTC AGAAGCAATA TTTTTTGTTA GCATCTTCAA	2580
ACCACCCTTT GCCGCGGCAT AAGCACTAAC AGTCTCACGA CCTAATTCAC TCATCATTGA	2640
GCAAATATTA ATAAATTTAC CCCCCTCTT CTCGATCATC TCAGGAATAA CCGCTTTTGA	2700
CATAATAAAC GGCGCATTCA AATCAACATC GATGACTTGT CGGAAATCTG TCGCTGACAT	2760
ATCTGTCAAT GGAATTCGTT TTATGATTCC AGCATTATTA ACTAAATAT CGATTGAACC	2820
AACCTCTTCT TTGATTGCT TGACCATTCG TTGAACTGCT TTTTCATCCG TTACGTCACA	2880

GACATAGCCT TTTGCTTCAA TGCCTTCTGC TTTGTAATTT GCCGATCCTT CATCGACTGA	2940
TTCTTGAGTC AAATTATTGA AAACGATCGT TGCTCCTGCT GCCGCTAATG ATTTGGCGAT	3000
TTCAAAGCCA ATTCCATAAA CTGCGCCCGT AATCAATGCG ACTTTCCCAT CTAAGCGAAA	3060
CATATCCATA TTAAAATCCA TCTATTCTCC CACTTTCATT CGTTCTATTT TAGTTCATCC	3120
ATTGGAACCA TATCCATGTC GGTATAAGTG ATATTCTCTC CACACATCGC CCAGATAAAT	3180
GAATAGTTGC TTGTTCTTAC ACCTGAATGG ATCGACCAAC TCGGTGAAAT CACGGCTTGT	3240
TCATTCCCCA TTACCAGATG CTTTGTCTCA GCTGGGTCAC CCATCATATG GAAAATTTTT	3300
GTATCTTCCT CCATATCAAA GTAAACATAC GCTTCCATTG GACGTTTCATG CGTATGACAC	3360
GGCATGGTAT TCCAAGAACT GCCCGGCTCT AAAATCGTAT AGCCCATTTG CAGCTGACAG	3420
CTTTCACAAA TATTGGGATG AATATATTGA TAAATTTTAC GTTCATTTAA TGTCAATGCT	3480
TCCCCAGTTT CCATTGGCTT AATCTCATCA ATACTGATTT TTACATTTGG GTACTTGTGA	3540
TGCGCCGGAG CAGAACTAAT GTAAAATTC GCAGGATCAG CTGCATCTTC TGAAGAGAAA	3600
ATCACTTGTC GAGTTTCTTT CCCGACATAA TAGCCATCTT GCTTCTTCAT CGCTTCTTTT	3660
CGACCGTCGA TTTCAATAAA ACCCGGACCG CCAATATTAA TGACTCCTAA TTCACGCCGT	3720
TCCAAGAAAT AATCGACACC TAATTCTTTG TTCAAAATAA TCTCTAACGA CTCAGTCGTC	3780
GGTGTTACAC CGCCAAAGAT CATGCGATCA TTGTGGGTAT ACGTCAAAC GACTTCATTA	3840
GGTACAAATA CTTTTTCAAC TAAAAATTCT TTCCTCAATT CTTCTGTTGA ATAATGACTA	3900
ATATCTTTTG GACTGTGCGT ATATCTCGTA TCCATATTCT GCATGTTTAT TTCCCTCTTT	3960
TACTTAATAT TTAACGAACG ATTGACACG AACCTGCTGT CAAAAATTGT TCGATCTCTA	4020
TTTCTGAAAA TTGATTACAA TCCCCGTGAA CAGTATGCTT CAATGCTGAT GCAGCCGTAG	4080
CAAATTCAAC GGTCTTTTGC CCGTCAAAAT CTGAGATCAA TCCATGCAAG ATCCCCGCCG	4140
CAAAAGCATC ACCTCCGCCA ACTCGATCAA CGATTGGATT TATTTTATGA ACTTCAGACT	4200
CATAATACGT ATTCTCCAGC CATAAAGTCC CTACCAACTG ATTGTTGCTA GCCGAAGAAA	4260
CGATCCTTTT CGTTGAATAG AAGCATTAA TATTGGGAAA CTGCTGCTGC ATTTGTTGAT	4320
AATAATAAAC GAATCCTGAA TTCATCCGGT CTTCTGGAAT CTCCAATAAA TAACGCGCAT	4380
CCAATTCTCC AGCTGAACAA TAATCGACAT AAGGCAAGAT TTCTTTTAAT GCTTTCCCAG	4440
CTGCTGCCTG ACTCCATAGT TTCCCACGAT AATTAATATC AAAGCTGACT AACACACCAA	4500
ATTTCTTTGC GGCTTTCCTT AATTCAACCG TCCATTGACA CCATTGCACT GACAAAGCTG	4560
CTGTAATGCC TGAAATATGA AATAGCTCAA TCCCAGAAAA AAGCTTTTCA TAAGACCACG	4620
GAAACTCTGT TACTTCCGCA AAGCTTGACC CCGCTCGATC ATAAGTTACA CTAGCAGCAC	4680
GTTACCCGAT CCCCCTTTCC ATGTAATAAG TTCCTAAACG CGCTCCGCCA AGTTGTACAA	4740
AATCACTCAT CACACCGTTT CGACGCAAAT GTTTCCTGCTG CGCTTGACCT AGTGAGTTAT	4800
CAGGCACTTT ACTTGCAAAA GAAACCTGAT GACCAAAATT GGCTAATGAA ATAGCTACAT	4860

TTGCCTCGCC GCCGCCATAG TGACAGATCA ATTGCTGTGT TTCGTTAAGA CGAACACCTA	4920
TATCCGTTGA AAGTCGCAAC ATAATTTTCGC CTAAAGTTAG TATGTTTGCC ATAATTAGTT	4980
TTCTTTTATT TCTCTAAATT TGGCCATATA TTGTTTTGCC AAAGCCGTTA CTTTACCAAA	5040
CTCACCAGTT TTAGCGTGAG CTAATAAGTT TCCGCCAACA CCCACTGCCG TCACACCCGC	5100
CTCAAACCAT GACGCCATAT TATCCAGACT AACACCGCCT GTGGGCATAA TATTCAAGTA	5160
CGGCAAAGGT GCTTGGAATG CAGAGATAAT ACTCGGACCA TAGGTACTTC CCGGAAATAA	5220
CTTCACGATA TCTACTCCGC TCTTCAAAGC TGTCTGCATC TCAGTAATTG TCATACATCC	5280
TGGCAAGTAA GGAAC TTGAT ACAAGTTACA AATTTTCAGCT GTCTCCCGAT TAAAGCTAGG	5340
GCTTACAATA AATTCGGCTC CGGCCATTAT AGCTAGGCGA GCCGTCACAG CATCTAACAC	5400
CGTGCCCGCG CCAATCACGA TATCTTTCTC GTCATGATAT GTTCAACTA ATGACCGAAT	5460
GACTTCATCT GCTTGAGGGA CTGTAAAAGT CAATTCTAAT CCAATCATTC CGCCCTCTAC	5520
TACTGCATGA CTAATTTTCA AAGCTTCCTC TTTAGAATCT GCTCGCAATA CTGCGATAAC	5580
TCCCGCATTT TCTAAGCGTT TCAATACTTC AACTTTTTTC ATTCGTCTAC GCCTTTCTTA	5640
TAAAGCTTTA TGGAATCAAA TATACGAGCA CCAGATTGAC TATACCGCGC TACTTCTTTT	5700
TAAATTCCTT CTTCTGGCAG TTCTTCTGTT tCATTGTTTT TTTCAAAaG CmTAATATtG	5760
CTCTTTCCAG CCTGTACGAC TGTTTCAGCA ATTTCAAATG AATCCGTGAC GCTATTTCGC	5820
ATCATCAAGC CTTCTAATAG TATTCCTAAG TTTGTGCCAC TCACAACAAA AACTTGATCT	5880
TTTTCTCAG AGAATTCAAC ACAGGTTTTA AATGGGGTAC CACCCAACAA ATCTGTGAAA	5940
CAAATAGCTT GATCAAACAT TTGAATTTCT GCCGCTACTT TTTGGCGATA CTCATCTACC	6000
GTCTCCTCGG CTAAAAAGGG AATAATGATC ACATCTTCTA GTTCTCCAGC AATCATCGCT	6060
AAAGCACTCT GCATCCCAAT CGAAAATTCA CTATGCCAG TAATTAATAA TTTCATCTCA	6120
TTGCTCCTTT ATTTAATCCG AATCTTTGCT TCCTTTTGAT CAATCAAAAC AAGCCGACCA	6180
TAAAAAAATT GGCCATATTC ACTCACTAAT AGTTTATTCC CTTTGACAAT ATCTTCCCGC	6240
ATAACAAAGA GATCAAATTT TTCCGATCCG GCGATTATCC GAAATCCCTT AGCCTGACAG	6300
AATTGCTCAT TTTCTCCTGT TTGACTCACT TTTATTGGCG TAATTGGAT ATCCTCTAAC	6360
GGAGAAATGA CTGTTGCTTG AATTTCTTTT CCCGTTTTGT AACAAAACCTT GTTTGATAAG	6420
CGTGGGTGCT CATTGAGCTG GTTATAAATC TCTGAACCTT TTGCCACTGA TTGCTGTGTT	6480
TGGCCTCCAG CAAAAAGGAG TGTGTACTTA TGCTTGTTTCG TAGTTAAAGC AAACCGATGC	6540
GCTTCTTTTT GACAATTTAT CGACGGTGCC AAATTATAGG TACTCGTAAT TTCAGTCTCT	6600
TTCTGTCCTG CAAAGCTATC AATAATAACT ACGGAGTTGA TCGACTTTAA ATAGATGAAG	6660
cTGCGCTCAA AAATCATTGG ATTCTGATCC GCCTTATCCA GCCAGCCGCA TTCTGCAAAA	6720
AAACCGACAG AAAGTTCTTT TATTTGCTGA AATAAGGGTG TCGGTAATTT GTCATAACCC	6780
CACGTATCGG AACTAAAGT ATGGGGATTT TCTGCGATAA ACATCGTATT GTGCGAAgCG	6840

CACTCTTTTA GCTGAAGTCG CTCTGATTG TTGACATAGC TGTAACGACC ACTATCGGAA	6900
AATAAGTCAT CCCCTTGTAG TTGTAGTGTA AATCCACCTG TAGAAGCATG ACCATGTGCG	6960
CTCCCATGCA GACCGTTAAA AAGTGTA AAA TAGATATCCT CTGCTTTGTA CGCCATCAGC	7020
CCACTTGATT CGCCACGAAA AAGTTCTTTT GGCTTCATTG TTTCCAGAT TCTTTCTTCA	7080
TAAAGATCCC CCGTCCAAAG CCTTGCCATA TTCGCAGTCA TGGAGGGTTC AAATATGAAA	7140
CCTAATTTGC GATAGATATC ATATACGTAA TGAAAGTTGA CATGATCACT ATCATTGATC	7200
GGATTTAGGA TATCTTG GTT ATCCGCCAAA TAATGGGTAG AGAAAATAGG TGTTTTTAAT	7260
TTCATGCGAA GATCTAATGG TAACTGGACT TCAAGATATT CAGAAATCTG CAATAGATAC	7320
ACGAATGTCA TCAAACTTC GTGCTGGTAC AGCGGGCTCT GCTCCCAATG AATTCCATCT	7380
GAATAGAATT GTAGATCAAG CTGTTAGCT AAACGAGACC ATATTAGATC CCTCTGTTTA	7440
CTGGTCACCA GTTCTGGAAG AAATAAATCA ATAGCTGCCA TTCCACCAAT TGCCAACACA	7500
CCCCAATTAC TGAGCCTGTA TTTATCGATA TAGGACCGCT CCAAATAGTC CAGATGGATC	7560
AGCAAGGCGT TGTTCAACAC ATCATCAATT CCTAATAGTC TGAAATCAGC GATTGGAATA	7620
TACGTCAAGC TTTTCATCCA GTTTGTTACT CGAATCCCAA CATCTAACGG ACGCCAAACA	7680
TCCCTATTCG TCGAATTTGG CTCACCCTCA TCGTTAATAA AATCAATAAG CAAGCTGTGC	7740
CATTCTGTA AGTAACGTTC TTTTTAGTA AGTGCATATG CTTGTGCCAG ATCTACAAGA	7800
AAGCTTTGAC GACTCAACAT AAAGAGCCAT TCTGGATCAT CATCAGGATA CCGATTCCAC	7860
GCATACTCTT TTAAAGAATA AGGAATGGAG CAAGCTTCCA TATCCATTGC ATCATTGTAA	7920
ATAATCTGAT CTTTCAGCAA ATAAGAAGTA CGTTCAAAAA GTGGCTTCAA AAATTGTGAC	7980
TGTTGCTCCA AGTACTCTGA AAGCCAATCT TGTTGAAAAA ATTTTAGTTG GTAGTCTTCG	8040
ATTCTTTGAA GATTCATCGT GTTCCCGCTT TCAAATTTTT ATTCGCCTCA GGCTTAAAT	8100
TACAGTATAC TTCCTCAATC ACTCCGAAAA GTGACAAATA CGGAGACAAG AATAATTTTC	8160
CAAAAATTAA AAATCTTGTT TCTGATTACC CTCGTTTACT ATGCTAAAAAT TCCTAGGACA	8220
GACGCTGCTA TCCCAATAGC AAATAGTAAG ATCAATAGTT TATACGTGTT CCAATTTCTC	8280
TTCTTGATCA AATAAAACAC TAAGATAGTT AAAATTACTG GAAGCATTTT AGGCATAATC	8340
TTATCTAAAA TATCCTGAAT AGCGATTACT TGTTTTTCAC CGGATTCTAC GTGATGCGAA	8400
TACTTTATGG CGATGTTGGC CTTAACAAAT GATGTTGCTA ATGCCGCTAT AACAGTTACT	8460
CCGACAATAT TAGCAGCCGA AGAAATTTTC CCGATCTTAT CACTCAATGT TTCGATGACA	8520
CTCGTTCCAA GTTTAAAGCC TAAATAGCCC ATTAATAACT TGATAATCAA CATCGAGATT	8580
AGCATTGAAA ACCAAAAGCC CATCGGTGCA AAACCCAAAC CATCCATTGC TAACCCGGCA	8640
AAGATTGTTG AGAATAAAGG AGCAATCCCA AATTGAGCGA TTGAGTCTCC AATTCCCGAT	8700
AATGGTCCCA TTAGCGCCAT TTTGATGCTG CGCGTATCCG TGACACTTTG TCCATTGTCA	8760
TACATTGCCA ACTGCATACT AGTAATGAAT GGCAGCATTT GAGGGTTGGT ATTATAAAAC	8820



TCTAAGTTTG	CACTCGCCGT	TTCTTTCAGC	TTCTCTGGTT	CATCTTTATA	GATTTTCTTC	8880
AGCGCTGGCA	TGATCGTGTA	TAAATAACTT	ACCCCTTGAT	AAGATGTATA	ATTAAAAGCG	8940
TTCTGTAACA	AATACGACCG	TAGTGTGTC	TTCATGTAGT	CTTTTGGCT	TAATTTTGGT	9000
GTCCTAGATT	CCATTACTAA	AATCCTCCTC	TTCCATCAAG	GTGGTTTGGC	TTTCGTCTCT	9060
TATGCTCTTT	TCTCCAACAA	TCGTTTTATC	TTTTCCATAA	TATTGAATCA	ACGCAAAGAC	9120
AGTTCCTACC	ATTGCGACAG	CCATCGTAGG	CAGCTCTAAA	TAAGCTGCAC	AAATATAGCC	9180
TAGTAACACA	TAAGGTGTCA	GACTCTTTTC	GAGCATCACA	GACATGATCA	TTGCAAAACC	9240
TATCGCTGGA	ATAAGCCCGC	CAGCAACAGA	AAAGCCATTG	ATTAGCCAAT	TTGGTAACGC	9300
TTCAACAAAT	GCCTGCAATC	CAGGACGACT	CAAGGCCGCT	AATAGTCCAA	TCATAAATCC	9360
TGCAACGGCA	AAAGCTATAT	ATGTTGAATT	AGCAATCAGA	TTAAAACGTG	TATATTTCCC	9420
TTGCTCAATT	ACATTCTTCG	ACCATTTAAC	ATTCCCAGAA	AATATAGTAT	AGATCGCCGT	9480
TGTTACCAGT	TGGAACAATA	CTGCAAAGGG	AAATGATAGT	GCTAACGCGG	ATTCAGGTGT	9540
TACCCCTTGA	TGTTTAAGCG	TAATGGCCAT	AATCGTACCA	ACGATTCCGG	GTCCAACCTG	9600
ATTGGGTGGT	ACGTTTCCGC	CTGCTCCAAC	ACCAAATCCC	ATATAAGCCA	ACTCGGCTAA	9660
AGCACCAAAT	GTTAAAGCCG	TTGGAATGTC	ACCCAAAACA	ACACCTACAA	AAAAGGACAT	9720
AACCAAGTGA	CGGTTTGTGT	AAATTCCTAA	TAGCTGCCCC	GaATAACAAA	AAGCTGTGAT	9780
CAATCCGATC	AATATACCTT	GAACAAATGA	AATCTCCATC	TGCTTCTCCT	TCTCCTTCTC	9840
CTTCTAATTT	AATTAGATAT	TTTTTCCATC	AATACATCTA	AGTACTGTGA	GCCATCATTT	9900
GAAAGTGGTG	AAGTCTTGGT	GTTGAAAGTT	ACACCATAGT	TATCTCGCAT	CAAGCAGAGA	9960
GCTTGCTTAT	CTTCTTTACC	TAAATAGATA	AATTGAGAGA	TTTTCTCCTT	GCCTTCGTCA	10020
GCATGAATAT	TCCCAATATT	TAAGTCTCC	ATTGGAAAGC	CCAATTTGCA	TAGTTCTAAG	10080
GCATCATGCA	AGTTTCCAAC	GACTACAAAA	ATTGTTTGAC	TAGGGGCTGC	TTTCCAAATA	10140
ACTTTTGCCG	TACGTTTCAT	CGTCCAAAAA	CGAATATTTG	TTCTTTTCGG	GACTACCGTT	10200
TTCATCAGCG	TCTGCTGTAA	GCTATCCTCT	GCCGCTTTAT	CATTGGCACA	AATAACTAAG	10260
TTCACTCCTA	AACTCTTGAT	CCATAGTTGA	CCTTGTCAT	GTATCAGACG	TTCATCTACA	10320
CGCACCATTT	TTACATTGGG	TTTTTGCAAT	TTCTTTCCTC	CTGTTTTTAC	CAGTACATTT	10380
TCCAATTTTT	TTTGAATCGC	ACTAAAGCTT	CTAGATAGAA	ATAATCACCC	CAAATATTAC	10440
CTTCATTTAC	GCCTTTACCA	CTATGCCAAG	AATAAACACC	CTCATTCAGT	AAAGCTGTAA	10500
TTCCGTCTGT	TTTGGATTCA	GTATAGTTTT	CAGATAAGCT	TCTTAGCATG	GAATGctGCG	10560
CACACTTGTA	AACCAACTTA	TGACTGTTTCG	TCTCAGGTAA	AAACGCGTCC	ATTAGACTCA	10620
TACCACAAAC	AGCAATCGct	GTCGCAGAAG	TATCTCGGGA	CTGTTTCGGAA	CCATCCATAA	10680
AAATCAGATC	CCAATAACTA	ACAAAATCCT	TCGGcAAGCG	ATTCAAGaAA	TAATTCGTAA	10740
TCGCCtCAAA	AATTTCTGTA	TTCTTTCCT	CTGGACAGTA	GGAGCGGAAC	AATGCCAAAC	10800

CATaAATCAA CCACGATTGT CCCCTCGCCC AGCTTGAACAT ATCGGAATAT CCTTGGCGGG 10860  
 TTTTTCATA CAAAGGATCG CCCGTTTCAG CATCGAaGAA AAACGTATGG TAAGCAGATG 10920  
 AATCTTCTCT AATCGCATT TCGATTGTCG TATGATAATG TTTAACGGCT ACTTCACGAT 10980  
 AACGTTTATC TTGAGTCACT GTTGACGCCC AAAATAATAA TGGAATATTA AGTAAGCAGT 11040  
 CTACGATCAA ACGATAGTTA TCTTGATTGC CCAATTCTCC CCAAGCTTGG ATAAATTCTC 11100  
 CCTTAGATTG GTAGCGCTGC AATAATTTTT CTGCTGCACT TATCGCGGCA TTCTTAGCAG 11160  
 CTTCAATTCC AGTCAACTTA TACGCACTTA CACAGGATAA GCTGTATAAG AATCCCAAAT 11220  
 CATGATGCTC GACTTCAATT TCATTTTCAA TTCGGTAAAT AAAACTCTTG ACATTTTGT 11280  
 CTGCAACTTC ACGGTAATGC TTTTCCCCAG AAACCTCATA GCACAGCCAT AATATTCCTG 11340  
 TCCAAAACC GTTTGTCCAT TCTGTATTAT CCATAGCTAT ATACGTAWTG TCCGTTGkTG 11400  
 ckGGTGTTGG GAATTaTCTC CCAGTTT 11427

## (2) INFORMATION FOR SEQ ID NO: 166:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1669 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

CCCTGnGACT ATTTTCTTA TCTCTTTTAT AAATAGAAGA GAAAGTATTC TGATTTTGTG 60  
 TTTTAAGCTT TTTATGCTAG AATAACGCCA GAACATAAGG AAAGGGACGA GwCAwTTGGA 120  
 ACCCATTTAT TTAGATCATG CGGCAACAAC ACCGCTTCAT CCAACGGTCA TTCAAGCGAT 180  
 GACTGAAAT ATGCAAACGA CTTTGGCAA TCCATCAAGT ATTCATCAGT TTGGACGAAA 240  
 GGCACACGGT CTTCTGGAAG AAGTGCGTCA AACGATTGCC GAGAGCTTAC AAGCCAAACC 300  
 CCACGAAATT ATTTTAAATA GTGGTGGCAC AGAAGGGGAT AACACAGCCA TCTTGGCAGT 360  
 TGCTTTTTCT CGTCAAAAGG AAGGAAAACA TATTATTACT ACGGCAATTG AACATCCCGC 420  
 TGTGTTACGA ACGATGGAAT ACTTAGAAAC GTTAGGCTTT GAAGTGAATT ACTTGCCTGT 480  
 GAATGaAAAT GGTCAAGATT CAATGGATCA ATTCAAAAAG TCTTTACGCG AAGAAACGAT 540  
 CTTGGTTTCA ATGATGTATG GCAATAATGA AATTGGAAAT CGATTACCGA TTGCTGAAGT 600  
 TGGTGCAATT CTAAAAATC ATTCGGCGAT TTTTCATACA GATGCCGTCC AAGCTTATGG 660  
 AAGTGAAGTC ATTTTACCTC ATGAATTAGG GATCGACTTA TTAAGTATTT CCGCTCATAA 720  
 AATCAATGGT CCAAAGGCG TAGGTTTTTT ATTTAAAGC GATGCAATCC AATTACCACC 780  
 TCTTTTACAT GGTGGGGAAC AAGAAGAAA ACGACGTGCG GGCACGGAAA ACTTAGCTGG 840  
 AATTATTGGC ATGGGTACTG CCGTTTCATT ACTAACTTCT GCAGAAAAGC AAGCAAGAAA 900  
 AACAGCCTAT CAAAGTTTTC AAACGATTAT TTTAAAGCG TTAGAAGAAG CCAATATTGA 960

TTTTtCmATT AATGGkGaAC CAACCAATCG TTTAGCGCAT GTCTTAAayC TTCATTTAAA	1020
nGGAATcCCA GTGATTTGCT CTTAATGCAT TTGGACTTAA GAGGTATTGC GATTTCAACA	1080
GGCTCGGCTT GTACAGCGGG AACTGTGGAT CCTTCGCACG TGTGACAGC GATGTATGGA	1140
GAAAATTCCT CAGCTATTAA AGAATCAATC CGGATTAGTT TTGGCTACGG GAACACACCA	1200
GAAGAAATTG CGACTTTTTT TGAAGTACTA GTTGCACTGA TTCAACAATT GAAAAAATAA	1260
TCTGATTGCG GCAGCTACAG GTTTTCGCTA TAATGAAGAT AACAAATGAAA ATAATGAGGT	1320
GGAACCCATG GCATTTACAA CAACGGCTGC TGTGGAAGGC TCAACAGTTT TTTACAAAGT	1380
ACATCCTAGT GCAAAACGCT ATACATTAAA AGATAACGGC TTTACTGAAA CAAAATCAGG	1440
GAACCTTCAA TTGATTGCTT CCTTAGATCC AACGCCTCAG AGAAATGAAG GCTTCAAATT	1500
GAAAATTACG ATCACTGCTG ATCTAAAAGA GTTGAAAATG TCCATTACAA CGGCAAATGG	1560
TTTAAAACCA ATGAATATTT TcAAGAATGA ACAACATGAA ATGAGTAAAG AAAAATACTT	1620
TTTCTTAATG GATGGCTTAA TTAGCCGTGG TGTATTAGAG AAAGTAGAA	1669

(2) INFORMATION FOR SEQ ID NO: 167:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 19031 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

GGTGAGTGTT CTAGCAGTGC TGTTTGTGGC AGCTGCTTGT TCCTCAAAAC CAGAAGCGAA	60
AATTAATACA GAACCATACT CTGACCGTCA AACAAATGCTT GGTACCTACG TTCAAATTCG	120
CATCTACGAT GATGGCAAAG AAGACGTCTT ACCGAAAGCC TTTGCCCGTG TCAAAGAACT	180
AGGCGATAAA ATCACCCTCA ACCAACCAGG TTCAGAAATT GATGAATCA ACCAAGAAGC	240
AGGCGTGAAA CCAGTCAAAG TCTCAGATGA CTTGTATCCT TTGTTGAAAA AAGCCTATGA	300
ATACAGCAAA GACTCCCGCG GCGGCTTTGA CATGGCAATT GGCCCAATCA CCTCTATGTG	360
GCACATTGGC TTCGATGATG CCCGCAAACC AAGCCAAGCC GAAATTGATC AAGCCTTGAA	420
ACTTGTCGAT TACACGAAAG TCAAATTCAA CGACAAAGAG CAAACCGTTT ATCTTGAAGA	480
AAAAGGCATG CAACTAGATT TAGGCGCAAT TGCCAAAGGC TTCATCACCG ATGAAGTCGT	540
GAAAGTTCTG AAAGACAACG GCGTTACCAC CGCTATTGTC GATTTAGGTG GGAATGTTA	600
CGTCTTAGGA CACAGCCCGC GTGGCAAAGA CATGGATTGG ACAGTCGGGA TTCAAGATCC	660
GAATAAAGCC CGCAACACAG TCTTAGGTCA AGTCAAAGAA AGCAACAAAA CCTTAGTCAC	720
ATCAGGAATT TATGAACGCT ACTTAAAAGT CGATGGCAAA ACCTACCACC ATCTCTTTGA	780
TCGTGAAACT GGCTATCCGT TTGATAACGA CATCGCGGGT GTCACCATTA TCACCGACAA	840
ATCCATCGAT GGTGATGGTC TCTCAACAGC CGTCTTCTCA ATGGGCGTCA AAAAAGGACT	900
AGAATATGTC GAAAGTCTGA AAGGAACCGA TGCCATTTTC GTCACAAAAG ATGACAAAGT	960

CTACATCAGC AAAGATATCG AAGGCAACTT TGAAATCGGC AAAGACTCTG GCTATACGAT	1020
GGGGAATCGG GCCGATTGA AATAAGGAGC GACTAAAAAT GTCTTTAAAT GTGTTTTTAA	1080
AAGTAGTGA AATTCAGACG AAGCTTGCCA GTCTCTTTCC CTTCAATTATC GGGGTCTTGT	1140
TTTCAATGAC CTATTTTCGGG GAAGTTCAGT GGGGCAACAC GCTGGTTTTTTC TTCATTGGCA	1200
TGGTGGTCTT TGATATGGCC ACTACCGCCA TTAATAATTA CATGGACTTC AAAAAAGCCA	1260
AATCAGACGT CTATAAATAT GAAGAAAATA TTATCGGACA ATCAGGCGTT TCACCACAAC	1320
TGGTCCGCAA CATGATTTTC GCCATGATTG CCTTCACTGC CGTTGTGGGT GCGTACTTGA	1380
CGGTTCAAAC CGGCTGGCTC TTCCTTGTTT TAGGCGGGAT TTGCTGCTTC ATCGGGATT	1440
TCTACACGTA TGGACCGATT CCGTTATCTC GCATGCCCTT GGGGGAAATT TTCAGTGGCT	1500
TCACGATGGG CTTAGGCATC TTCGTCTTAA CCGTTTACTT GAACGTCGTC ACCAATCCGC	1560
CTTTTTACTT AACGGTCGAT TTTGCTAGCG GCGCCTTTCG CTTAGACGGC AATCTCTGGG	1620
CGGTGCTTGC GATTGTCGTG GCGTCCTTGC CCTTGGTCTG TTCGATTGCT AACATCATGC	1680
TCGCCAACAA TTTACGGGAC TTAGACACGG ATATTGAGAA CCACCGTTAT ACATTAGTCT	1740
ATTATATTGG GCGCCCATCG GGGATTGTGC TGTTTCAAGT GTTAGCGCTA sctGGGTATC	1800
TGGTGATTTT ATTCGGCTTC ATTAGTGGTA TTTATCAATG GCCTGTGTTA ATTACATTTG	1860
CCACTTTACC TGTCATTTGG AAAAATATTC AAACGTTTAA ACAAGAATTG CCGCAGCCGA	1920
AAAGTTTTCG CCATTCCATT AAGAACTTAA TGGTTTTCAA TGGCACCTAT GCTTTAGGGC	1980
TATTCATTAG CGCCTTACTT GGTAGTCTT GCGAAAAAAT AGCGGGCTGG GCGAAAAAAT	2040
TACTGTTTTT TTTCAGAGGC CTGCGTCAAA ACAATCAAAC AAAGCTCAAA ATGTTAGTAT	2100
TCGTTGCTAA CATTTTGAGC TTTTCTTGTT AAAAAGAGCG GAATCTTAA ATTTTGTGAG	2160
GGGATTTTGG TAAAAATTTG CCGAAATCGC CCTTTTTTTG AGCCGATTAT TAAATAACAG	2220
CTGTTGTAAC TCTTGTCAGA ATAGGAACCT TCAAAATTTT CGCATGGGGT GCGAAAAAAT	2280
CGTCAATTTT TGCTGGGGGT TTCTTTGTTA TTATAGAGGA GTAACAAAAG GTTGTTTTGT	2340
CTTTCTTGTTG TAAAAGGGCA AGAAAGGCTA GCGAGTTAAA AGGAGGTTTT TCAATGAAAA	2400
AAAAGCGTTA TTTAATGATT GTGTGTCTAC TATCTTCTCC TAGTTTTTTT ATAAATGTTG	2460
AAGCGTCTGA TGGTGGTTCT AGTTCGGTGG GGATTGAATT TTACCAAAT CCGAGAACAC	2520
CCGCTCCTAA AGATCCCCCA CCGAAAACAG ATGCGCCAGC TGCTGATCCC AAGGAACCAG	2580
CTGGTCCTCC GCAAGGAGAT CAACGAAgTG GTGGTTCGAC ACAGACCACC ACAACTGGCT	2640
CAACGCTCCC TCGTACAGGG AGCAAGAGTC AGGCAAATTT GAGCATTCtC nGnTTCGCCT	2700
TAATCGGTTT GGCGGGAATC GTACATAGAA AGAAGGGACG ACATGAAGCA AACTAAGTGG	2760
CAACGATTAG CAACCATTGG CTTGTGTAGT TCTTTAGTAA TTAACGCCTT TTCTGGTGTG	2820
ACGGCAGTTG CGGAAACCGT GACGATTGAA AGTAGTCCGA CCGCCGAAAG TAGTGCCAAG	2880
GAAGAGACGC AAGCAAGTAG CGTGAAGGAA GAAACAACGA AAGCCAGTAC GGAAAATAGT	2940

CAAGTAACAA CTGACACGAG TCAGGAAGAA GCAACGAAAG AAGCGGAGAA AGAAGAACCG	3000
CAAGCAGAAG TGGAACAAGC AGAAACACCA ATCATTCTTA AACCAAAAAA AATCAATATG	3060
AAGGCAACTT ATTCATTTTC TGCAGAACT TATCAGTTTG GATTTGTGAA TGAATCAGGT	3120
CAATTAATAA ATCCAGATAT TATACCAATT ACGTATAGCT ATGCCAAAGG ATCATGGAAG	3180
ACAGATGGTT ATAATCGAAA GTGGACTAGT ATGGTTCAAG GGAGTGCTTC AACCGTAGGA	3240
AACCTAAAGA ATGTAATAAT GCCAGCAACT TCTGTAGTTA TGCCACCAGG ACCGTCATAT	3300
GAAGGAACTC AAGAGGTGTA CACAACTTT TCAATTCGCA TACCAAAATA TTATGCATCA	3360
GCGAGTCTCT ACAATAGAGA AGGTAAAATT GATTCTACTT ATCCGTTACC TGCTATTGCA	3420
CTAGCAGGTA CTAGACCGCT ATCTTTGACT CAAAGTAGTG TAATTAGTGC ATTGGCGCTG	3480
ACCAAGTAAAG GAGACAATGT TTATACACCA CGGGAAACAT TTTTGGAGG AGATCCTGCA	3540
GGTGTAAGT TTACTAATTT TTTGTATCGT ATAAATGACT TTGATGTGAA AGGTAATAAC	3600
ATAGGTTATA AGACTGTGAG TAGCCCAATC TATTACCATC TGACCAACCG CCGTGTCAAC	3660
GAAACTTCG TAGATACAAG TGGCGCCAAA ATCACGCCAC CAAGTAATTT CACCCAAGGG	3720
AAACAAACGG TCATTAACAG TGATCCTTAC ACGTTCCAAC AAAGTGGTTT TTTACCCGAG	3780
ACCTACAAAG TTGGCACGAA ATCTTACCGA TTCAAAGGCT GGTACAAAGG GAAAACCAAA	3840
ACCGAGCCTT TGGCCACCAC TAAACACCT AGCTATAAAG TCACGTATGA TGACAATGAT	3900
GATTTGACGG TGGTCTATGA GGAGTTTCA GGGTACGAGC TGCCTGCTTC GACCAATCAA	3960
TTTGCTTTG TGGATGAAGC GACGAACAAA TTAATTGCCC CCGACCAAGT GCAGATGAAG	4020
TATAATCTTA CTTTAAATGA AAATAATAAA AAAACAGTAA TGAGCAGTAA CTTAACGGGG	4080
ACAGATACAG CGACACTGAA AAAGTTGTCC GTGCCTGTCA ACTATTTTGA ACAATATCGC	4140
GTCAATACGT TTTATGGCGC GAGTGACATT ACGTTTACAT TGCCCAAACG GTACAAATCA	4200
ATCAATATTA CCAAATCAGA TGGCAAAACC GACCCAGCTT TTCCTCTTCC TAAATCTAT	4260
AATATAGATC AAGTAGAAAT GTCACACATG CCTGTGACCA CTTATAACAA GTTGAAACAG	4320
CTGTCGGGCC AAACGTTTGG CTTTAATGCT TTAGCCGATC AACCTGAATT TTATACGAAA	4380
ACGTTATTTG GGACAGAGTC TGGCATCGAT GACCCAGTCA ATTATTATAC AATGAGTGGC	4440
CCTGTTTACT ATTATTTAGA AAACCGCAAA GTCACCGAGA aCTTCGTAGA CACCAACGGC	4500
GCTAAATCA CACCGCCAAC AGGTTTCACC CAAGGTAAAA AAACGGTGAT TACAAGCGAC	4560
GCCTACACTT TCAAACAAGC AGGCACCTTA CCAGACACTT ACACAACAGG CGGTAAGACC	4620
TACAAGTTCA AAGGTTGGTA CAAAGGCAAG TCCATACTCA ACACATTGAC AACTACCAAA	4680
GCGCCAAGTT ATCAAGTGAC CTACGATGAC AATGATGATT TGAATGTGGT GTATGAAGAA	4740
GAAACAGTTA CGACAGTGTA TCCATCAGTC GATATGAACT TTGTGAATGA AAAAGGCGGG	4800
GCTTTCACAC CGGCGTTAAC TTTTAGTGGT AAGTACTATG CGCAAAGTAC GAGTGCGTAC	4860
TTAAGAACCG ATTTATATGA CGTGACCTCA AAAAATAATG GTAATGGGCA ATATACGGTA	4920

AGTATTAATA ATGGTAGTAT GCCATTGTCC CAAGAATTAT TGAAAAAATA TAATAATGGA	4980
CAACCAATCA GTGCTACCAA CAGATTACAG TTTAATGTTG ATAAATTAGC CATCGACCAA	5040
CAACTAAAAT ATGTTGACAG CATTCAATTA GACACAGCTC AAAGTAGCAA TCTGAAATCC	5100
TATAGATATG TGTACACGAA CAATAGCTCA CTGGTTTTCG ACCCAAATGT AGCACCAGCA	5160
GAGGTTGACC TTAGTTCAGA ATCTCTTAAC TTGCTTAATT TTGATTCAGA TGGCACCTAT	5220
TTTTCTAATG CAAATAATAG ACTTTTTTAC ACGCATTAG GATATAGTGG CACACCAGGA	5280
GTTAACTATC TTCTCGTAAT GTTCTTTTT AACGCCAAAc CTGCGGATAA GTCAAACTT	5340
GTCTACAAAG TCACTCGCAA ACAAGTCACC GAAAACCTCG TGGATGTCAA CGGTGCCAAA	5400
ATCACTGCAC CAACAGGCTT CACCCAAGGT AACCAAGTAC CAATGAACAG TAACACCTTC	5460
AAGTACACAG CGGCAAAAGC TTTACCAGCG ACGTATACTA CAGGTGGCAA AGTCTATACG	5520
TTCCAAGGGT GGTATAAAGG GAAAACCAAG CCAAGTACGT TGAACAAAAC AACAACTCCA	5580
ACGTTCAATG CGACCTTTGA TGGCAATGAC GATATGACCG CCATGTATAA GGAAGAAATA	5640
CCAACAGCTA GTGTCACATT AACTCGACCA AAAGAAGTGA TTGATACGAA TACCAATGTA	5700
ATCTGGACAA CAACGATCAC GAATACTAGC AAAGCACCTT TACAAAATCT CACCTTGAAA	5760
AAAGGGCCCA ATTGGTCAGC TGGTCTGACG ATCCCGACCT TTATGGAAGT GACACCAGAA	5820
GGAGAAACGA CAAAaTCAAT CCCAGTAAAT AGTACACTTT GGACAGAGGG GGTTCCTTTA	5880
CCAAATGCCG TTCcTATCGG CAAAAAGTT TCAGTTGCTT TCACAACTCG CGCAACAGGG	5940
AAACCAAACA CTGTTTTGAA AGCAGAAGTT GTAGTATTTG GTGGTATTAA AGATAGTACA	6000
GTGGATAACT TCGTGAGAAT TCGTCCAAAT GATCAAGAAG TAGTCACACC AACGACCGAA	6060
GGCTTCATCA GTGTGCCAAC CTTCGACTTC GGCCAAGTGG GCGTTGCAGG AACTAAGCAA	6120
TAACACAGCT TGAAACAAGC CGCGGATTAC TACGGTAACG GCACACGGAA TCCGTATCTG	6180
CGGATTAAGA AAACGCAACC CAATTGGAGC TTAACAGCGC AACTGTCACA ACCAAAATCA	6240
GCGACAGACA GCTTGCTTAC AGCGACCCGC TTATTATTAG GGGCGGCGCC TGTCTCTAGC	6300
TTTACCAATT ACAATCAACC AACCGAGTTG AAAAATACGG TCGGTACCAC GAGTGCCATT	6360
AGCTTAACAG CCAACAACAC AGCAACGAGT ATTATTGCCA ACAAGCAATT CACAGGTAGT	6420
AATGTTTATC AGTTGGAATT CACCTTCAAT AATGTCAAAC TTGAAGTGCC AGCCAATCAA	6480
GGTGTTAAAG GGCAACAATA CAAGGCCGCA GTTACATGGA ACCTAGTTAC AGGTCCTTAA	6540
AATAGTTAAG TTATCATGAG AAAAGCTCAA AATGTTAGTG GTTATCACTA ACATTTTGAG	6600
CyTTTTGGTT TAAATATTAA GGAAAAACA GTTGGAGTTC ATCAAAAGAG CGTTTAAATC	6660
CTTGATTTTC GCTAAAATAA CACTTATTTT CAAAGGGAAC AGAGGCGCTC AGCCTTGAAA	6720
GGATAGGAAT CACTAAAAAA CGGCACTCAA TATGTCAAAA TTTGAAATTT CAAGCTGTGT	6780
GTTCTTTGGT AAAATAwATA wAAAAATGCT AGTTATCAGT ATCGATAATA ACAGGATACT	6840
GATTAAGAAA GGACTTTATA GAGACTATAG ATTGAATTTT TACATAGAAA GAAGGAGCAA	6900

GATGAAGCGA	GTAAATTGGA	AAAGATGGCT	AGTTGTTGGG	TTAAGTTGTT	CTTTGTTTCAT	6960
GGATTCAGTG	GTTGGTGTGA	CTGTGTTAGC	GGAAACGATT	ACTGGGGCGA	CGGAGCAAGG	7020
AGTAGCAACA	TCTCAGTCGA	GTGACGAAgC	GAGCCAGACG	aCGCAAACAA	CCGaAGAGTC	7080
ACAGGCAACG	GTCGCTAGTG	AAGCGAAAAC	AGTACCGCCA	CAGGAAACGG	CAAGAATTGC	7140
TTCTCGAGCG	ATTGGTTATT	CTTCTGTGGA	AGGGCGCGAG	ATTCCCTTTT	TCTTTGTGGA	7200
GGAAGACGGG	ACGTTGTTTG	ATCCCGACCG	AATTACGATG	GCGGTCAATC	TTTCCACGTT	7260
TTCGTTTTAT	GAAGAGAAAT	TACAACGAAC	CCCCCTTGAG	CCCACCACTG	TGAATGGCGG	7320
AAAGTTACTG	TCTATTCCAA	CGTCACCAGC	TTTTAAATAT	GATACAAATA	ACCAGAATCC	7380
AAGTAATATT	TATGGCGTTT	CTGAAGTGTC	GTTTACTATT	CCTAAGGAGT	ATCAAAGCCT	7440
GGACATTTCGA	CCAAGTACGT	TTTATACAGG	AGACACTACG	CAATATCCAG	TGCCAACGGT	7500
TTTTGCGAAC	GTTGGGGGCA	AAGTGACGAA	CTATGTGGGC	GCCAATGCGG	AGACGGAATT	7560
AGAGTTAACC	AATGAAAAAA	TGCCCAATAA	GCTGACGTTT	GGTCCTAAAA	AGACGTTTAA	7620
ATATACGGTA	GCTACGGCAC	CAGGAGGCGT	TACGTATGCG	CTGACCTATT	TTTATGGAGA	7680
TGTCGGCGGT	CCAACTAGTT	CGCACCAAAG	ACGAGGAACA	GCGGGTCCTG	TGTATTATTA	7740
TTTAACAAAG	CGGCGTGTCA	CGGAAAAATT	TGAGAATCCC	GCAGGCGGGG	CGATTCTGTC	7800
GCCAGAAGGT	TATACGCAGG	ATAAGAAAAC	CATTGTAACA	GGGGAGGATT	TTACTTTTAC	7860
CCAAGAAGGC	ACCTTGCCCTG	AACGTTACAC	AGGCAGTGAT	GGGAAGACGT	ATTTATTTAA	7920
AGGTTGGTAC	AAAGGGAATG	CGAAACCTAG	CACGTTGGA	ACCACCAAAA	CGCCTAGTTA	7980
TGCGGTGACC	TATGATGACA	ATGACGATTT	GCATGTGGTC	TATGAAGAAG	CAGTGATGAA	8040
AACCTATACG	TTGCCAGCGA	GAGAAGCTTT	GTTGCGCTAT	GTTGATGAGC	AAGGAACTT	8100
GATTAATCCC	GCCAAGTTTA	AGCTAAGTGC	GACCATGGGT	GAAAGTGACG	GAGCCACAGG	8160
GGAAATGACG	ACTTTTCCCA	CAATTGATGG	AATCGATATG	CCAGCAAGTC	AATTAAAGAA	8220
ATTAGCCATC	CCGCAAAAAG	TCTACACACG	CCCAGACGAT	GGGACAATCG	TAACTTATGG	8280
CCCGCAAGAA	GTGAGTGTTG	AAATTCTTAA	GTATTACCAG	ACGATTTCGA	TTTCACCAAC	8340
TACTGCGTAT	ACAGGGGATA	AAACCAAGTA	TCCAGTACCA	AATGAAGTGC	GCCGTGGCAT	8400
CGAAAACCCC	GACAACATTG	TTAGTAGTTT	aGTGGGAaMc	mCTGCGTATA	ACTTGACCCA	8460
AAAAAGTGCC	ACACGCTATA	CTGCCC GCCG	TTCTTACTGG	AmGTGGGGCC	CCACGAAGAC	8520
ACTTTACTCA	ATGAGTATCT	ATTCAGGAAC	TGCTGGGGGC	AACTATAATT	TATCGACCCC	8580
TGATGGCACC	ATTTATTATT	ACTTAGAAAA	TCGGCGGGTC	ACTGAACATT	TTGTAGACGA	8640
AAGTGGCGCA	AAAATCACGC	CACCAACTGG	CTTTACACAA	GGAAATCAGC	TAGTGGTGGA	8700
CAGTGA AAC	TATGTCTACA	CTGTGCAAA	AGCTTTGCCG	AAGATCTACC	AAGCTGGTGA	8760
AAAAACCTAT	ATCTTCCAAG	GCTGGTTTAA	AGGCCAAAACC	AAGCCAGCAA	CATTAAAGAC	8820
GACAACGACC	CCAAGTTTTA	CACCAACTTT	TAATGATGAG	GACGACATGA	CCGCTGTGTA	8880

CCAAGAAGCG ATTCCCACCG CGGaAcTAAC GTTAACAGGT GCCGTTGACA TAATCGAAAA	9940
TGGCGCCACA ATGGATTACT GGGAGGCGCT ACTGAAGAAC ACAGGCGAAg cGCCGTTAAC	9000
CACCATTAAA ATCAAGCCAA CGGCAACTTG GCGGCTGGC ATCGGCGCAC CCAACACGAT	9060
ATTTGTACAA GGAACGGGTC AAAACACCAA AGCTTTTCCT GTCACCAAAG AACAATGGAC	9120
GACCGGTGCA GGAGTGTCCA TCACGTTGGA TCAGCCTTTA CCAGCTGGCG GTCAATTAAA	9180
AATGAACTTA TTAGGAACCG CCGTTACAGG AAATCCTGGT CAAGTTTTAA CCGCTGATGT	9240
TGAAGTAACG GGCAACTTTG GCAGTTTAAC TGCCAAAGAT ACGGTCCGTA TTAAAGACTT	9300
AGATCAAGAA ATTACGAGTC CTGACGGCGA CGGCTTTATT AGTACCCCGA CATTTGATTT	9360
TGGTAAACTA GCAATTTTCAG GAAGTAAGCA ACAATATGGT TTGAAGAAGG CCGCAGATTA	9420
CTACGGCAAT GGCACCTCGA ACCCTTATTT ACGCCTGAAT ACTAGCCAAG CCAATTGGAG	9480
TTTAACGGCC CAGcTATCGC AACCAAAATC AGCCACAGAC AGCTTGCCAA CAACGACCCG	9540
CTTGTTGCTA GGAACGGCCG CTGCTGCCAG CTTTACCGAT TACAACCAAC CAACAGAAAC	9600
CAGGACACCA CTTGGCAAGA CCAGCACCGT GACTTTAACC GCCGACAATA CCGCAACAGC	9660
GGTGGTCGCA AACCAACAGT TCACAGGCAG TGACGTCTAT CAGTTGGACT TCACGTTTGC	9720
TAACATCAAA CTAGAAGTGC CAGCCAACCA AGGTATGGCT GGCCAACAAT ACCAAGCCGC	9780
CGTCACGTGG AATTTAGTGA CTGGCCCCTA AAATAAAAAT GCCCGAGCCT TAGAAAAATA	9840
AGAGGCTCGG GCATTTTTAT TGATGGTCTT TTGATGGTCG TTGTGCTATA ATAGCATTGA	9900
AAAGAGAGAT ATGCTGTAAC ATACCTCTCT AGTGTAGAGC CGTTTAAGAC GGTGACCGAT	9960
TTTGTTACAA AAAATAACCG TACTCAGTCA AAGTAAACGG TTATTTTTTC TTGTCTTCTT	10020
TAACAACCTT TAGGATTCCG AAGATTAAAG TCGCGATAAA AGAACCGAAA CTAATCATCA	10080
GAGCTAAAGC TTCTGCGATA GACAAAAGGC CTCTCCTTTC ATAGATTTTA GTACTTACAT	10140
TCATAAGCAC CACCACCTTT CAGATTGGAT AGCCACCGCC ATAACCTCTC TACACTTTTA	10200
TTATAACCATT TAGAACTAAA AAATAATTAG TTTTCAATAT ATTTTTTTAG GAAATTGTGC	10260
TATAATAGCA ATGAAAAGAG AGATATGCTT CAACATACCT CTCTGATGTA GAGCCGTTTA	10320
AGACGGTGAC CTTTTTAGTT ACAAAAATA ACCGTACTCA GTCAAAGTTG ACGGTTATTT	10380
TTTATTGTCA TTTTAAAGCT TTAAAAGATT TTCCTGTTCC CTAATTGTTA AGAATAGAGA	10440
AGAAAAGCTC AAAATGTTAG TTCTGATTAC TAACATTTTG AGTTTTTTTC AATAAATATC	10500
AAGAAAAATC AGCGGAAATT AGAGAGAAAA TGGTGAATTC CACTGATTTT CAGGCTGTTT	10560
TTTACCAAAA AACAGTGGCT CCTCCTTGGT ATAAAAGGAA TCATTGGAAA TATCAAGTTG	10620
TCGTATCAAA ATTTAAAATT TCAGTGGGGC CTTTTTTTGA TAAAATGAAA ATCATACAAA	10680
AAATCATACA TAGGACAATA AAAATAACAA TAAAAAGTTA CATAAAAAAG GAGAATGGCA	10740
TGAATCAAAA AAGAAACCGT TTATTTGCAA TCGCTTTGTG TGGTTCTTTA CTAATCAGTT	10800
CATTAAATGG CATCCCAGCA ATTGCTGAAA CCATGGCGGC CCCAGATGTT CAAGGGGAGG	10860



ACCGCCAAAC AGTAGTCCCG TCAGAGGGGG TTCAAGGACG GCGCGCCGTT TCTTTTCCAG	10920
AGGCAGATAC GGAAGTnAGc GTACCGCCAG CTGATGAAGA GCTACCAGAA CAGACTGAAC	10980
AAGTAGTCGG GCAAGTTGAA GCGTCAGTGA CGAGTCCAAT AATAGTTAAT ACACCAGAAT	11040
CACTGCGTGC GCAgCGAGCA GTGCAGACAT TCGATATTCC AGAGAAGACG TACAAAATGT	11100
CGTTTTTAGA TGAAAGTGGT CGTTTTATTG ATCCAGCGAA AGTTTCGATT ACTGGGGATA	11160
TTATGAAGTT TACTATGCAA GGCCTTTGA CGAAAGTCGG GGCCGTGGCC ACTAATAATG	11220
TGGGAACCAC GAAGCAATTG ACGGTGCCGA AAGTAGCTTT GTCTGGGGAT ACAGCCATAA	11280
AGAGTTATGA TTTTCCGGC ATCCAAAATG AAAAGATTAC TTTGCCGCGG TTCTATAATG	11340
TTCCGAAAAT CACGCCAGT GCGAATTACA CTGGACAAAT CTATCCAATG GCAGTGACGA	11400
AACAGGTGAA TGTTACAGGA GAAGTATCGA ATGAGGAAAC CACAGGCGCG AGGCTGAATA	11460
TGTTTGTATC AGGTATCCCT AACCAGTTTG TCTTTTCTCG TGCTAGTTGG GACAATGATC	11520
CAGCGGGAAA TATGTTTGCG CGTGGCTTTA ACATGGCTAC TGCAGGCGGC ACACCAGCTT	11580
ATCTGACTAC GGATGATCAC ATTTACTATT ATGTGCCTAA CAAACGGGTT AACATTTATT	11640
ATAATGACAT GACCCCGAAT GGCTTACCAG CGTATCCTCC AGGGTATAAT GCCAGTCTTT	11700
CAAAAACAGT AGTTGATTCA GAAAATTTTC ACTATAAAGC GCCAAAAGCT TTTCCAGAAT	11760
TCTTTAGTTC TGGGAATCGT CACTTCCAGT TTAAAGGCTG GTATAAAGGG CCTGCAAGAC	11820
CAGCGGATCC AAAAGCCGTT AAGTTAGAAA CCTCATTGAC GCCAGAGTTC GATGTCACTT	11880
ACGATGGTGC CGATAATTTA TATGTTTTTT ATGATGAAGT GAAAGAAATG ACCACTACTA	11940
TTCCAGAAGT TACCTATAAG TTTGGCTTTG TTGATGAAAA AGGGGCGCTT GTTGCGCCAA	12000
CGAATGTAGA CATTACGGCG AACTTAACAA CTGTGAGGA TAAGGTGGTC CAGAAGGTGG	12060
GGGCTATTAC AGGCGCTAAT GTAGGCAATT TAAAACAGTT AACC GTTCCT AGCCAAACAC	12120
TAACCTATGT ACCTAAAATT AAAGTTGCTT CTTCTGGGGT AACC GATTTT CGGTTGACTA	12180
TTCCCAAGCG CTATAAAATT CCAACAGTCA CACCTGGTTC ATTCTACACT GGCTCAACGA	12240
CCGCTATCC TTTGGCGACA ACGCTGTAA GGCATATTAG TGGTCAGGCA GATGAACGGA	12300
TAACAACGGA TGGCACTCGC TATAGTTTAT ATAATTTACC GACGCCTCAT AGCTATCGTA	12360
TGTACAGAAG TAGTTGGCAA CCTGACGTTA CCAACACGGG ATTTAGTGCA ACTTTAAACA	12420
TGGCGACGAC CGAGATTTAT CCAGCTTATT TCACGACCGA TAACACGATG TACTATTTCT	12480
TAGAAAACCG CCGCGTCACT GAACACTATG TTAATGAAGC AGGTGCTGAG GTACCGATGC	12540
CTACAGGCTT CACGCAAGGG AATCAAACGG TGATTGACAG TGATACTTAT CATTTTAAAT	12600
TGGCAAAGGA ATTGCCGTTT AGTTATTTTT TAAACAATAA AGCATACCGA TTTAAAGGCT	12660
GGTATAAAGG CAAAACCAAA CCTAAGGTAT TGGAAACGAG TCGAACGCCT GAGTATGATA	12720
CAACTTTCGA TGATAACGAT GATTTAACCG TTGTTTACGA AGAAATTAAA TACGGCGGAA	12780
ATGCGGTGAC ATTTGGTTTT GTTGGGGAAG ATGGGCAATT GCTTCAGCCG ACAGGGTTTC	12840

AGGTGACTAC	GGATATTGTT	GAAACAATCG	ATGGGCTGAG	TACGGTTTTA	AGTAATGTTT	12900
CAGCTGTCGA	CAGTACAGGC	AACGTTAAAA	CTTTGACGAT	CCCGCAAAAG	GAATATGTAT	12960
TGACTCCCCA	GATGAATTTT	TATGGAACAA	AAAATTCATT	GATTACGATT	CCAAGACAAT	13020
ACCAAGAAAT	TTCTTTTACC	AAACCTGCCA	ACTACCAAGG	ACTAGATTAT	CCAGTTGCAG	13080
GGCGTGTTGA	GAATATGTTT	AGTGGTAATC	TGTATGTTGA	GGGGGCACAA	CCACCTTACT	13140
TTTCTGCTTC	GAAAGTCAAA	GCCAATCAAT	ACAAAATAAT	TGAGTTTCAT	TGGAAAGAAG	13200
AGCCTACACA	AAGTTTCTCT	AGTTTGTACA	AAGCTACGTT	AATTCGAAGT	AGCGTGGCAC	13260
CAACAGGCGT	AGAAATGTAT	CCTAATAAGC	CAATCTATTA	CTATGCGACC	AATCGCCGAG	13320
TGACTGAAAA	CTTTGTAGAT	AAATCTGGCG	CAAAAATCAC	GCCACCTCAA	GGATTTACGC	13380
AAGGGAATCA	AGTCACAGTC	ACCAGTGATC	CTTTTACTTA	TACTGCTTCA	AAAGCTTTAC	13440
CAAGCGTTTA	TGCGGCTGGG	GCAAAAACCT	ATAAGTTTGT	TGGTTGGTAT	AAAGGAACCA	13500
CCAAGCCAAC	GACCTTAAAA	ACAACGGCTA	CTCCTAATTA	TTTGGTGGAT	TTTGATGATA	13560
ATGACGATAT	GACGGCGGTC	TATGAAGAAG	AAACGCCGAC	TGCTGCTTTA	ACTTTGACCA	13620
GCACCAACCG	TGTCGTGAAT	AATGACGATA	GTGTAGATTG	GGTGGCAACA	CTGAAAAATA	13680
CTAGCCTGGC	ACCACTTAAA	ACGCTGACAG	TGAAACCAGC	GACTACTTGG	CCTGCAGGAA	13740
TTGGTACGCC	GACGTCACTT	TCTGTCCAAT	TGGACGGACA	AGCACCGAAA	ATTTATCCAG	13800
TAACGGCGAC	TACTTGGTCT	GAAGGAATCA	GCTTaACAGG	TCTTGAAATA	CCGGCAGGTC	13860
AAACAGCCAA	CGTGACGCTT	GTTGGCACCA	AAATTTCAGG	AACATCAGAC	CAACGCTTGA	13920
CGGCAACCCT	TGATGTTACA	GGAAATTTTG	CGACAGTGAG	CGCCGCGGAT	GCGGTACGTC	13980
TGACAGATAC	AACCCAAGGA	ACAATTACAC	CAACAGAAGG	CTTTATCAGC	GTGCCAACCT	14040
TTGATTTTGG	AAAAATGAAT	ATTGCTAGTA	AAACGCAACA	ATCAGGCTTG	AAAAAAGCGG	14100
CCGATTACTA	CGAAAATGGT	ACACGCAATC	CTTACTTGCG	AATTAAGAAA	AATCAACCGA	14160
ATTGGCAGAT	GACAGCGCAA	CTATCTCAAC	CGAAAGCAAC	CACAGATAGT	CTGCCAACAG	14220
CGACTCGTTT	ATTGTTAGGT	CCAGCGAATG	TTTCAAGTTT	TACAAATTAT	AACGAAGCAA	14280
CCGAACAAAT	CAAAGCAGTT	GGTAAGACTA	GTTCTTTGAG	CTTAACGGCG	AATAATGTGG	14340
CAACCAAGTG	TGTCGCAAAAC	CAACAGTTCA	CAGGTAGTGA	CGTCTACCAA	CTGGATTTTA	14400
CGTTTGAGAA	CATTAAATTA	GAAGTACCAG	CCAACCAAGG	AACCAAGGA	CAACAGTACA	14460
ATGCTGCGGT	CACTTGGAAT	TTGGTGACTG	GTCCCTAAAA	TAAAAATGCC	CGAGCCTTGA	14520
AATAGGAGGC	TCAGGCATTT	TTTATTGATG	GTCTTTTTAG	GGTTGTTGTG	CTATAATGAT	14580
AAAGAAAAGA	GAGATATGCT	TCCAACATAT	CTCTCTGATG	TAGAGCCGTT	TAAGACGGTG	14640
ACCAATTTTA	TTATTTAAAA	ATAACCGTGC	TTGGTCAAAG	TAGACGGTTA	TTTTTTCTTG	14700
TCATTTTTAA	GCAATTTTAC	AATCAGCGCA	ATCAAAGCAA	TGGTAAACAT	ACCAAACCCA	14760
AGAATTGTCT	GAATTGTTTC	ATATGCTGAC	AAAAGGCCTC	TCCTTTCTGT	GAATTTCGGA	14820